

UNIVERSITÀ DI PISA



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Corso di Laurea Triennale in Ingegneria Informatica

Generazione di mappe di crisi dinamiche e interattive

Tesi di Laurea

Candidato:

Gabriele Ara

Relatori:

Prof. Marco Avvenuti

Prof. Maurizio Tesconi

Ing. Stefano Cresci

Ing. Fabio Del Vigna

Anno Accademico 2015/2016

Sommario

In caso di eventi catastrofici come terremoti o attacchi terroristici, l'organizzazione dei soccorsi richiede l'accesso al maggior numero di informazioni possibili in tempi assai ridotti. In questo senso, l'attività di *Social Sensing* permette di ottenere una grande mole di dati dai contenuti pubblicati dagli utenti dei Social Network, molto prima che le notizie relative a questi eventi possano essere diffuse attraverso canali ufficiali. Perché questi dati possano essere sfruttati al meglio da parte di organizzazioni umanitarie come la Protezione Civile, è necessario però che da questi dati sia possibile estrarre le informazioni chiave in modo chiaro e immediato.

La produzione di mappe di crisi a partire da set di dati così ottenuti, permette di realizzare una presentazione chiara ed efficace delle informazioni ricavate. Questo, però, può non essere sufficiente, qualora tali mappe siano realizzate in modo statico.

In questo lavoro descriviamo la realizzazione di un sistema per la generazione automatica di mappe di crisi dinamiche e interattive a partire da set di dati strutturati, ottenibili tramite analisi di messaggi raccolti da un Social Network. Le mappe prodotte da questo sistema permettono una esplorazione dei dati che evidenzia la loro evoluzione temporale su più livelli di dettaglio, permettendo di ricavare un numero di informazioni di gran lunga superiore rispetto a quelle che potrebbero essere ottenute tramite soluzioni statiche.

In seguito, viene presentato un caso di studio che permette di illustrare il funzionamento del sistema prodotto con dati reali, ottenuti a partire dai *tweet* inviati nella prima ora successiva alla scossa principale del Terremoto del Centro Italia del 2016.

Infine vengono esposti alcuni margini di miglioramento applicabili al sistema presentato, che potrebbero permettere ulteriori forme di visualizzazione dei dati registrati o la possibilità di monitorare l'evolversi di situazioni di crisi anche in tempo reale.

Indice

1	Introduzione	7
2	Related Work	9
2.1	Raccolta ed elaborazione dei dati	9
2.1.1	Analisi semantica e tagging dei messaggi	10
2.1.2	Twitter come fonte di informazioni	11
2.2	Crisis mapping	11
2.2.1	Tipologie di mappe tematiche in letteratura	12
2.3	Realizzazione di mappe coropletiche interattive e animate	17
2.3.1	Normalizzazione	17
2.3.2	Classificazione dei dati	17
2.3.3	Animazione di una mappa	18
3	Lavoro Svolto	20
3.1	Introduzione al sistema sviluppato	20
3.1.1	Struttura dettagliata del sistema	21
3.1.2	Uso del sistema	22
3.1.3	Change Blindness e soluzioni adottate	22
3.2	Controllore della simulazione	23
3.2.1	Invio degli aggiornamenti	24
3.2.2	Controllo dell'esecuzione e cambio dei parametri	24
3.2.3	Cambio del set di dati	25
3.3	Livello coropletico	26
3.3.1	Normalizzazione	26
3.3.2	Classificazione dei dati	27
3.3.3	Zoom e cambio della visualizzazione	29
3.4	Prestazioni del sistema	30

<i>INDICE</i>	5
4 Caso di studio	32
4.1 Il terremoto del Centro Italia	32
4.2 Fasi del lavoro svolto	32
4.2.1 Raccolta e geoparsing dei dati	32
4.2.2 Realizzazione della mappa	33
4.3 Considerazioni sulle mappe ottenute	34
5 Conclusioni	35
5.1 Sviluppi futuri	35
Appendice A. Percezione dei Colori	38
Ringraziamenti	43

Elenco delle figure

2.1	Esempio di mappa a simboli proporzionali	13
2.2	Esempio di mappa ad area variabile	14
2.3	Esempio di mappa a densità di punti	15
2.4	Esempio di mappa coropletica	16
3.1	Schema UML del sistema	21
3.2	Mappa prodotta attraverso il sistema elaborato	23
3.3	Controller della simulazione	24
3.4	Menù di controllo	25
3.5	Cambio della mappa con lo zoom	30
3.6	Ricalcolo del set di dati corrente	31
4.1	Controllore dell'annotatore	33

1. Introduzione

In seguito a eventi catastrofici come terremoti, tsunami o attacchi terroristici, è indispensabile che informazioni sulle aree più colpite siano disponibili il più rapidamente possibile a coloro che devono organizzarne i soccorsi. La comunicazione di queste informazioni deve essere possibile anche mentre questi fenomeni sono ancora in corso, al fine di permettere un'organizzazione migliore dei mezzi di soccorso e del personale adeguato.

Per questo motivo, spesso si ricorre alla creazione di *mappe di crisi* [1], ovvero di mappe che evidenziano l'entità dei danni a cose e/o persone attraverso simboli grafici di immediata comprensione; queste mappe vengono usualmente prodotte partendo dai risultati dei sopralluoghi nelle zone colpite o di altre rilevazioni.

Spesso sono gli stessi enti di protezione civile a mettere a disposizione mappe riportanti valutazioni sull'entità dei danni. Le mappe attualmente prodotte dalla Protezione Civile italiana stimano l'entità dei danni in una data area a partire dall'intensità dei fenomeni fisici rilevati da certe stazioni fisse (ad esempio stazioni sismiche); queste stazioni però sono in numero assai limitato e quindi spesso non permettono di ottenere stime assai precise per ogni zona colpita da questi fenomeni [2].

Per aiutare la Protezione Civile a valutare con maggiore precisione la condizione delle aree più colpite, si cerca di ricorrere quindi ad altre tecniche di raccolta dati in tempo reale, che permettano di ottenere informazioni su aree anche scarsamente coperte dal raggio d'azione delle stazioni sismiche dislocate nel territorio.

I Social Media (SM) risultano particolarmente efficaci come mezzi di diffusione rapida di informazioni: in questo caso, ogni utente aggiunge informazioni potenzialmente utili ai soccorritori raccontando semplicemente la propria esperienza personale. Diversi studi hanno dimostrato l'importanza del ruolo che hanno avuto questi mezzi di comunicazione nell'organizzazione degli aiuti umanitari in eventi come, ad esempio, il terremoto di Haiti del 2010, lo tsunami del Giappone del 2011 e molti altri [3, 4, 5].

Visti in quest'ottica, gli utenti sono generalmente indicati come *Social Sensors*, cioè fonti dirette di informazioni relative all'ambiente che li circonda, anche se questi ne sono tipicamente

inconsapevoli. Secondo la US Geological Survey (USGS) [6] i principali vantaggi di questo meccanismo rispetto all'impiego di sensori "fisici" sono la rapidità nella raccolta dei dati e il basso costo di implementazione. Altri vantaggi portati da questo approccio sono quelli della pervasività nelle zone a rischio e della natura più variegata delle informazioni che vengono raccolte (ad esempio immagini, video, ecc.) rispetto a quelle che potrebbero essere ottenute dalla loro controparte "fisica". D'altro canto informazioni ottenute con questo metodo hanno lo svantaggio di non avere una forma tipicamente strutturata, poiché espresse in linguaggio naturale.

Una volta pubblicate, tutte queste informazioni possono essere raccolte da un sistema che permetta di analizzarle per trarne conclusioni preliminari su quali siano le aree più effettivamente colpite e la loro effettiva condizione. I soccorsi possono così organizzare le risorse a loro disposizione, che non sono infinite, in modo da ottimizzare il loro utilizzo e intervenire in modo più tempestivo laddove ce n'è più bisogno. Avere una chiara conoscenza della situazione del territorio e delle infrastrutture presenti permette di evitare inutili dispendi di energie e spesso di salvaguardare la vita stessa dei soccorritori, evitando loro inutili pericoli. La presentazione delle informazioni ottenute deve quindi essere il più possibile chiara e di immediata comprensione.

Le mappe di crisi prodotte in letteratura, sia a partire dai dati delle stazioni di rilevazione, sia da quelli ricavati da Social Network, sono affette dal problema della staticità: esse, infatti, rappresentano tipicamente una sorta di fotografia della situazione rilevata in un dato momento, che quindi non è in grado di mettere in evidenza come questa situazione evolva nel tempo [3, 7]. Per ovviare a questo problema si può pensare di affiancare più mappe di crisi relative a tempistiche differenti, lasciando al lettore la possibilità di confrontarle manualmente, attraverso ispezione visiva. Si tratta ovviamente di una soluzione poco efficiente: spesso, infatti, alcune particolari scelte in fatto di colori, scala o forma attribuita ai dati visualizzati può falsare l'interpretazione del lettore, per il quale può risultare assai faticoso distinguere le differenze tra le varie mappe.

La soluzione adottata in questo lavoro è stata quella di creare un sistema che, a partire da un set di dati contenenti informazioni geo-spaziali e temporali, permetta di creare una mappa di crisi interattiva e dinamica capace di riprodurre una simulazione dell'evoluzione dei suddetti dati nel tempo.

2. Related Work

L'approccio comunemente impiegato per la realizzazione di mappe di crisi consiste in un processo di raccolta ed elaborazione dati [2, 6, 8, 9] scomponibile in varie fasi:

1. Raccolta dei post o dei messaggi degli utenti di uno o più Social Network.
2. Elaborazione dei dati ottenuti; in questo caso le informazioni ricercate sono essenzialmente di due tipi:
 - (a) Presenza o meno di danni a cose o persone e (se presenti) in che entità [3, 9].
 - (b) Indicazioni di tipo geografico, come coordinate, menzioni di luoghi conosciuti, ecc.

Una volta individuate informazioni di questo tipo a partire dai dati raccolti, queste devono essere organizzate in un formato che favorisca il loro utilizzo in future computazioni.

3. Produzione di una presentazione efficiente delle informazioni strutturate ottenute al passo precedente.

In questo lavoro ci soffermeremo principalmente sull'ultima fase di quelle illustrate; dopo una breve digressione sulle fasi precedenti faremo quindi principalmente riferimento alla produzione e divulgazione vera e propria delle mappe di crisi a partire da un set di dati strutturati e organizzati.

2.1 Raccolta ed elaborazione dei dati

In questa prima parte del lavoro, vengono prelevate dalle banche dati dei Social Network tutte le informazioni relative a messaggi contenenti specifiche parole chiave o *hashtag* che possono riferirsi all'evento catastrofico in corso (esempi possono essere "terremoto", "tsunami", "attacco terroristico" e così via).

Per produrre delle mappe di crisi risulta necessario a questo punto poter associare ad ogni messaggio (o comunque al maggior numero possibile dei messaggi) delle informazioni geografiche, in modo da poter individuare a che luogo si riferisce il messaggio in questione. Purtroppo, dati statistici [9] dimostrano che solo circa l'1% dei messaggi totali presenti nelle banche dati dei principali Social Network possiede allegate delle informazioni geografiche sul luogo da cui è stato spedito. Queste informazioni allegate, inoltre, sono spesso un misto tra dati rilevati da dispositivi mobili al momento dell'invio del messaggio (tipicamente ottenuti tramite GPS) e la locazione di default specificata dall'utente nelle impostazioni del suo profilo.

Ovviamente, una percentuale così bassa di informazioni, in parte non accurate o non veritiere, non è sufficiente affinché analisi volte allo studio dei danni possano essere ritenute abbastanza accurate. In questo lavoro sarà mostrato come l'arricchimento dei dati attraverso annotatori semantici porti alla realizzazione di mappe di crisi più complete e precise.

2.1.1 Analisi semantica e tagging dei messaggi

Un approccio innovativo all'arricchimento dei dati con le coordinate geografiche consiste nell'esaminare il testo vero e proprio di ogni tweet alla ricerca di menzioni a luoghi particolari, attraverso l'uso di *annotatori semantici*, ovvero di strumenti software che permettono di associare al testo in analisi delle entrate di una *base di conoscenza*, in modo da associarne un significato [10].

Una base di conoscenza è un database contenente informazioni di carattere enciclopedico sotto forma di dati strutturati (e quindi facilmente accessibili attraverso strumenti software); esempi notevoli di basi di conoscenza pubbliche sono Wikipedia e DBpedia. I frammenti del testo che vengono collegati ad una pagina della base di conoscenza sono chiamati *spot*.

Di solito questo collegamento viene fatto con una certa probabilità di certezza, detta *confidenza*; ad alcune parole, infatti, possono essere associati più significati anche all'interno della base di conoscenza stessa. Per esempio, la parola "rombo" può indicare sia il pesce che la figura geometrica, a seconda del contesto.

Per eliminare problematiche di questo tipo, in seguito all'individuazione delle parole contenute viene spesso effettuata una opera di *disambiguazione* tra le varie entrate della base di conoscenza, col fine di eliminare il più possibile problemi come quello riportato. In genere, questa operazione dipende dal significato complessivo attribuito dall'annotatore al testo che si sta analizzando, detto *topic* o contesto.

Tornando all'esempio del "rombo", se l'annotatore riceve in ingresso la frase "Ho acquistato un rombo fresco al mercato" è molto probabile che deciderà di collegare "rombo" con la pagina di descrizione del pesce, assegnando quindi a questo collegamento una confidenza

relativamente alta. Agirebbe probabilmente in modo completamente diverso qualora la frase in ingresso fosse “Il rombo è una figura geometrica i cui quattro lati hanno la stessa lunghezza”.

2.1.2 Twitter come fonte di informazioni

Tra i vari Social Network, Twitter risulta per diversi motivi particolarmente idoneo come fonte di dati per questo tipo di studi [1]:

- Tutti i messaggi, chiamati in questo caso *tweet*, sono pubblici e il loro accesso è libero da restrizioni dovute a particolari impostazioni di privacy indicate dagli utenti (cosa che invece non è possibile con altri Social Network, come Facebook ecc.).
- Il limite di 140 caratteri per tweet impone agli utenti di focalizzarsi sugli aspetti essenziali dell’informazione che intendono trasmettere, facilitando non poco il compito degli annotatori alla ricerca di luoghi geografici o di altre informazioni in esso contenute, come ad esempio entità di danni a cose o persone.
- I tweet hanno tipicamente una struttura ben definita; i messaggi possono contenere quattro categorie di entità: hashtag (*#tag*), riferimenti ad altri utenti (*@utente*), URL e testo semplice. Queste entità possono essere sfruttate per filtrare la totalità dei messaggi inviati in una data finestra temporale secondo determinati criteri.

Inoltre, sempre [1] mette in evidenza come il numero dei messaggi relativi ad argomenti “di tendenza” su Twitter aumenti molto rapidamente: questo permette di ottenere in breve tempo grandi moli di dati che possono essere impiegati con lo scopo di creare mappe di crisi.

2.2 Crisis mapping

Sebbene molti possano pensare che l’attività di *crisis mapping* sia una branca relativamente recente della cartografia, la sua storia rivela che già nella seconda metà del ’600 Luigi XIV di Francia commissionò delle mappe di crisi col fine di organizzare coi suoi generali alcune manovre di guerra; mappe di questo tipo sono state largamente impiegate anche nelle due Guerre Mondiali [5].

Negli ultimi tempi comunque, è possibile affermare che la cartografia abbia subito una vera e propria rivoluzione, grazie alla capacità sempre maggiore di fare affidamento su sistemi di calcolo sempre più precisi e moli di dati sempre più grandi a costo ridotto. Il passaggio da mappe statiche su supporto cartaceo a versioni dinamiche e interattive richiede di rivoluzionare sempre di più il nostro modo di interagire con esse, esplorando modi sempre nuovi per

presentare dati georiferiti. Ad esempio, mappe animate permettono di comunicare in modo immediato cambiamenti nel tempo e nello spazio dei dati visualizzati [11].

In generale possiamo individuare più direzioni in cui è possibile portare innovazione nella visualizzazione dei dati:

- Studiando nuove tecniche di presentazione statica dei dati: questo significa accantonare per un momento tipologie di mappe preesistenti e introdurne di nuove che permettano esplorazioni dei dati differenti da quelle possibili in precedenza.
- Concependo nuovi modi di interagire con le mappe preesistenti o animando il loro contenuto: i dati visualizzati in una mappa possono cambiare nel tempo ed è possibile proporre altre visualizzazioni allegate in seguito ad eventi come il click su di una figura o similari.

In questa sezione saranno analizzate le tipologie esistenti di mappe per uso statistico, valutando quali possano essere considerate le più adatte per la realizzazione di mappe di crisi. Una volta individuata la più adatta ai nostri scopi, nella sezione successiva verranno analizzate le principali problematiche che sorgono quando si cerca di animare il suo contenuto nel tempo.

2.2.1 Tipologie di mappe tematiche in letteratura

Le mappe di crisi sono sostanzialmente delle mappe tematiche, perciò valuteremo in questa sezione quali tipologie di mappe tematiche esistono e le loro caratteristiche, per poi individuare quella più adatta ad essere impiegata per la realizzazione di una mappa di crisi¹.

Carte a simboli proporzionali

In queste mappe, il dato statistico da visualizzare è presentato in base alla dimensione di un particolare simbolo scelto per l'occasione, che può essere un simbolo geometrico (un cerchio o un poligono) oppure una figura che abbia un qualche collegamento col contesto della visualizzazione (ad esempio una automobile).

Un esempio di queste mappe è presentato nella Figura 2.1; in questo caso sono stati usati dei cerchi al cui interno è stato inserito anche un numero rappresentante la quantità del dato visualizzato.

¹Il contenuto di questa sezione è basato su quanto riportato da https://en.wikipedia.org/wiki/Thematic_map e da <http://indiemapper.com/app/learnmore.php?l=index>; in particolare, per il secondo ci si riferisce al contenuto della sezione “*Univariate Thematic Map Types*”.



Figura 2.1: *Mappa di crisi sviluppata durante la guerra civile libica. La mappa rappresenta la distribuzione degli aiuti umanitari delle Nazioni Unite durante la guerra civile nel 2011.*

Un problema comune di mappe di questo tipo è rappresentato dalla congestione/sovrapposizione dei simboli, soprattutto quando le aree cui si riferiscono non sono ben distanziate tra loro.

Inoltre, alcuni lettori possono incontrare delle difficoltà nell'interpretare in modo preciso il dato rappresentato, a causa della difficoltà di stima precisa dell'area di un simbolo. Molte persone sottostimano sistematicamente differenze tra aree e il grado con cui si tende a sottostimare queste differenze aumenta con l'aumentare della differenza di dimensioni tra due simboli adiacenti; conseguenza di questo problema è che spesso alcuni lettori tenderanno a sottovalutare l'entità dei dati rappresentati.

Carte ad area variabile

In questo caso il dato visualizzato è rappresentato sempre dalla dimensione di un simbolo, come nella tipologia precedente, ma la forma del simbolo stesso è data dalla forma geografica dell'area cui il dato si riferisce. Oltre alla superficie, anche la forma dell'area geografica di riferimento può subire delle deformazioni per enfatizzare ulteriormente il dato che si intende trasmettere.

Un esempio di una mappa di questo tipo è riportato in Figura 2.2, dove la dimensione e la forma degli stati rappresentati sono state deformate per rappresentare la loro densità di popolazione. Si noti come ad esempio l'Australia risulti assai piccola rispetto all'India o alla Cina, a causa delle sue vaste aree prive di insediamenti urbani.



Figura 2.2: *Una mappa ad area variabile. La forma e le dimensioni degli stati rispecchiano la loro densità di popolazione. Fonte: Indiemapper.*

Come per la categoria precedente, queste mappe sono affette dal potenziale problema della cattiva percezione del lettore di differenze tra le aree. Inoltre, qualora le aree non fossero sufficientemente regolari, come nell'esempio della figura, questo problema diventa ancora più accentuato dal fatto che la distorsione di una specifica area può essere male interpretata, qualora il lettore non abbia una precisa idea di come quella area dovrebbe essere in condizioni "normali".

In generale è quindi consigliabile usare queste mappe quando si vuole mostrare grosse disparità tra aree confinanti tra loro, senza troppe pretese di precisione, tipicamente contrapponendo tra loro un numero limitato di classi di dati.

Carte a densità di punti

Il queste mappe, il dato da visualizzare è rappresentato attraverso differenze nella distribuzione di punti o simboli nel territorio. Queste mappe sono particolarmente adatte per indicare in quali aree tendono a concentrarsi certe distribuzioni dei dati da visualizzare.

Un esempio di una mappa di questo tipo è riportato in Figura 2.3.

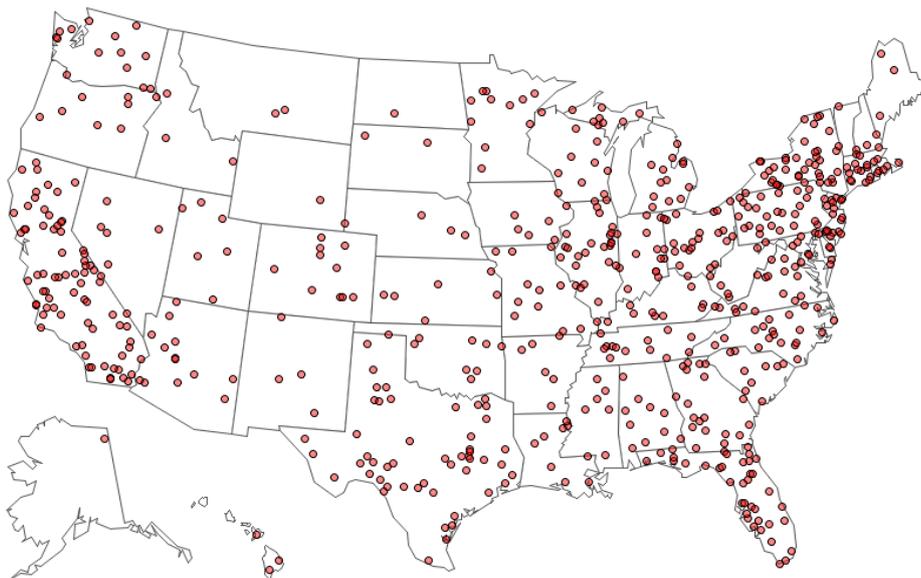


Figura 2.3: *Distribuzione della popolazione degli Stati Uniti, 2010. Un punto sulla mappa equivale a 500 000 persone.*

Sebbene il loro impatto qualitativo sia molto elevato, risulta ancora una volta difficile estrapolare dati precisi da queste mappe. Inoltre, il loro contenuto dipende fortemente dal tipo di proiezione che viene impiegata per realizzare la mappa stessa: la sfericità della Terra rende difficile la sua rappresentazione su superfici differenti da quella di un *geoide*; per visualizzarla è necessario definire una proiezione che stabilisca una corrispondenza biunivoca tra ogni punto sul globo e ogni punto sulla superficie della rappresentazione².

Le mappe a densità di punti richiedono che la proiezione impiegata per la creazione della mappa corrispondente sia una proiezione “Equal Area”, ovvero tale che zone con la stessa area nella mappa corrispondano zone con la stessa superficie sulla Terra. Molti software per la visualizzazione di mappe interattive usano una proiezione che non rispetta tale proprietà, perciò i dati mostrati da una carta a densità di punti realizzata con tali strumenti possono fornire una impressione distorta dei dati rappresentati³.

²Per visualizzare le differenze tra varie proiezioni, sono reperibili online vari strumenti che permettono di metterne diverse a confronto. Uno di questi è disponibile all’indirizzo <https://earth.nullschool.net/>.

³Per rendersi conto quanto queste differenze di aree possano diventare evidenti con la proiezione sbagliata è sufficiente aprire Google Maps, scorrere fino al livello di zoom minimo e confrontare un paio di stati con superfici note. Ad esempio, la superficie dell’Antartide apparirà molto superiore a quella della Russia, quando in realtà quella della Russia è di circa 16 377 742 Km² mentre quella dell’Antartide “solo” circa 14 000 000 Km²!

Carte Coropletiche

In queste mappe, ad ogni area viene associata una diversa sfumatura di un colore in proporzione al valore relativo che deve essere visualizzato. Il loro impiego richiede quindi in genere di suddividere lo spazio totale in zone e per ogni zona assegnare un set di dati da rappresentare; tipicamente si possono adottare confini territoriali come quelli di regioni o province, che sono ben definiti.

Questo comporta una certa aggregazione delle informazioni trasmesse al lettore, il che rischia di togliere un poco la specificità di certe informazioni geografiche, soprattutto rispetto alle mappe a densità di punti, riducendo però gli effetti di un eventuale rumore nei dati presentati. D'altro canto, mappe di questo tipo non sono affette dalla problematica del tipo di proiezione usata, visto che i confini dei dati rappresentati sono ben definiti.

Una mappa coropletica di esempio è riportata nella Figura 2.4.

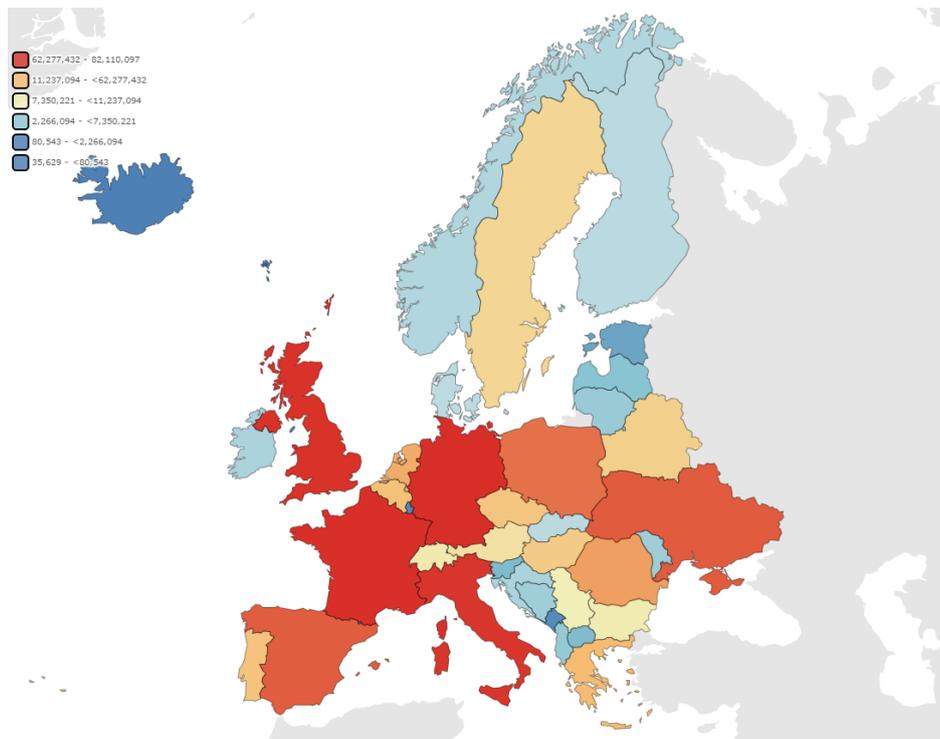


Figura 2.4: *Mappa coropletica raffigurante la popolazione totale dei singoli stati europei nel 2013.*

In seguito ci focalizzeremo su questa tipologia di mappe, poiché sono le più comunemente usate tra le mappe tematiche [12] e dunque anche le più documentate in letteratura, soprattutto dal punto di vista delle mappe interattive e dinamiche [13].

2.3 Realizzazione di mappe coropletiche interattive e animate

In questa sezione prenderemo in analisi le principali questioni che sorgono nel momento di voler realizzare un sistema per la creazione di mappe coropletiche interattive con contenuti variabili nel tempo.

2.3.1 Normalizzazione

Mappe di questo tipo mostrano i dati sulla base di una suddivisione dello spazio totale analizzato. Questo implica che per mostrare conteggi o una misure che dipendono fortemente dall'estensione della superficie considerata il risultato può variare anche di molto a seconda della suddivisione adottata⁴.

Questo si verifica spesso, nell'ipotesi in cui la distribuzione di un dato fenomeno sia uniforme tra varie aree. Supponendo, ad esempio, che la densità di popolazione sia relativamente uniforme in alcune regioni, noteremmo che il valore della popolazione totale in ogni regione è direttamente proporzionale all'estensione ad essa associata.

Per confrontare efficientemente aree di estensione differente tra loro (come regioni, province, ecc.), è dunque sempre consigliato ricorrere ad un meccanismo di normalizzazione dei dati, prima di rappresentarli sulla mappa. Ad esempio, invece di confrontare la popolazione totale delle singole regioni si potrebbe decidere di rappresentare invece la densità di popolazione delle stesse: in questo modo, il dato visualizzato non dipende più dalla superficie dell'area relativa, ma unicamente dalle caratteristiche intrinseche del territorio considerato.

2.3.2 Classificazione dei dati

Nella produzione di mappe coropletiche, e in generale di mappe tematiche, risulta di notevole importanza il problema della ripartizione in un certo numero di classi. È abbastanza noto in cartografia come uno stesso set di dati possa produrre un set di mappe assai diverse tra loro, a seconda dell'informazione che il creatore della mappa intende trasmettere al suo lettore [14].

È possibile rappresentare dati su mappe coropletiche senza doverli raggruppare in classi [15]. Questo approccio prevede in generale di "lasciar parlare i dati da soli", ovvero di assegnare ad ogni possibile sfumatura dei dati in ingresso una altrettanto sottile sfumatura dei colori sulla mappa. Sebbene questo approccio possa andare bene per permettere al lettore di apprezzare il più possibile la diversità dei dati raccolti tra le varie zone, rischia di non essere sufficiente per

⁴Vedi <http://www.esri.com/esri-news/arcuser/winter-2013/understanding-statistical-data-for-mapping-purposes>.

l'attività che ci siamo preposti nella creazione di mappe di crisi, ovvero focalizzare l'attenzione del lettore verso le aree più colpite, che quindi vanno adeguatamente evidenziate rispetto alle altre.

Inoltre, spesso stampanti o monitor tendono a esprimere malamente sottili differenze di colore come si potrebbero incontrare in dati non classificati ed effetti di contrasto simultaneo⁵ potrebbero ingannare facilmente l'occhio di molti lettori.

2.3.3 Animazione di una mappa

Come tecnica di presentazione di dati su mappe, l'uso di animazioni digitali è un'introduzione relativamente recente; al contrario di altre tecniche più collaudate, questa è senza dubbio quella in cui la letteratura si è avventurata di meno e il suo potenziale è spesso messo in discussione negli articoli di settore [13]. Il principale argomento di dibattito in questo campo è rappresentato dalla problematica di trasmettere correttamente al lettore i vari cambiamenti che avvengono in diverse zone della mappa; questo compito risulta relativamente complesso, a causa di un fenomeno caratteristico dell'occhio umano chiamato *Change Blindness*.

Il fenomeno di Change Blindness si verifica quando un individuo fallisce nell'individuare cambiamenti anche macroscopici che occorrono all'interno del suo campo visivo [16]. Si verifica spesso quando il lettore non presta la dovuta attenzione alle zone del suo campo visivo coinvolte in questi mutamenti e tipicamente questo avviene quando queste zone rientrano a far parte della sua visuale periferica, anche se ciò non è strettamente necessario.

Nelle mappe coropletiche animate, cambiamenti dei dati nel tempo sono rappresentati attraverso mutamenti nella tonalità, nella saturazione e nella luminosità dei colori che riempiono le sue aree. Poiché i dati rappresentati sono fortemente legati alla loro posizione geografica, ci si trova davanti alla limitazione di non poter presentare dove si vuole certe transizioni, che possono verificarsi in qualsiasi zona dello schermo, in base alla posizione della relativa area geografica di riferimento. Questo impedisce di attirare l'attenzione del lettore verso queste zone in modo più diretto, richiedendo l'adozione di tecniche particolari tali che le transizioni che si verificano non passino inosservate al lettore a prescindere dalla loro posizione nel suo campo visivo.

In particolare, cambiamenti nei valori visualizzati devono essere abbastanza sostanziali da permettere al lettore di individuare correttamente una differenza saliente rispetto ai cambiamenti che possono verificarsi ad esempio nelle informazioni di sfondo della mappa stessa [13]. Questo può essere considerato come un ulteriore fattore che spinga verso l'adozione nell'uso di

⁵Vedi l'Appendice A.

classi nella rappresentazione dei dati, poiché la rappresentazione di dati privi di un qualsiasi raggruppamento rischia di produrre cambiamenti troppo sottili perché il lettore sia capace individuarli correttamente.

Change Blindness e interattività

Un primo strumento considerato spesso come carta vincente nella creazione di mappe dinamiche è rappresentato dall'interattività della mappa in questione. Diversi studi [17, 18] confermano come strumenti che forniscano al lettore un maggiore controllo sulla presentazione dei dati (ad esempio la possibilità di eseguire le animazioni in loop, di scorrere l'animazione frame per frame, di modificare la velocità di riproduzione, ecc.) possano ridurre molto gli effetti negativi di Change Blindness.

Aggiungere transizioni tra i vari stati

Un'altra tecnica proposta in letteratura per evitare Change Blindness è l'uso di transizioni tra vari step temporali dell'animazione [19]. Studi in questo senso però riportano risultati contraddittori [20] e sembra (anche se non è ancora chiaro) che l'entità dei benefici/svantaggi apportati da questa tecnica dipenda dalla durata delle transizioni stesse [13].

3. Lavoro Svolto

In questa sezione viene presentato il sistema che è stato prodotto per la generazione di mappe dinamiche e interattive, a partire da un set di tweet già georiferiti, attraverso annotatori semantici o per mezzo di “tagging manuale”. Il suo impiego nel confronto di più annotatori semantici su dati di casi reali verrà presentato nella sezione successiva.

3.1 Introduzione al sistema sviluppato

Il sistema sviluppato si presenta come estensione della libreria open source per la visualizzazione di mappe interattive *Leaflet*¹. Leaflet garantisce un sistema per la creazione di mappe indipendente dal tipo di browser impiegato, rapido e leggero e che può essere impiegato anche per la creazione di siti web per dispositivi mobili. Uno dei punti di forza di questa libreria è la possibilità di estendere componenti già pronti messi a disposizione, in modo da poterli adattare alle proprie esigenze.

Questo sistema è composto principalmente di due estensioni tra loro parzialmente dipendenti:

- Un elemento di controllo, che permette all’utente di creare una simulazione di un evento e di intervenire sulla sua riproduzione, attraverso una interfaccia simile a quella di un player video.
- Un elemento di visualizzazione grafica, che crea un livello che si sovrappone alla mappa visualizzata e che mostra i dati relativi al tempo corrente della presentazione. Nel caso specifico, questo livello crea una mappa coropletica.

In particolare, la classe che implementa il livello contenente la mappa coropletica implementa una interfaccia compatibile con quella richiesta dal controllore della simulazione per la ricezione di aggiornamenti a proposito dei dati da visualizzare.

¹<http://leafletjs.com/>

3.1.1 Struttura dettagliata del sistema

Il diagramma UML delle classi del sistema è riportato in Figura 3.1. L'interfaccia `DataLayer` espone un solo metodo, che permette di impostare il set di dati da visualizzare; questo metodo accetta inoltre una funzione di naming, in modo da poter associare all'oggetto corrispondente nella visualizzazione i singoli dati contenuti nel set.

Come si può notare dallo schema, il sistema è stato progettato secondo il pattern *Observer*; infatti, molti oggetti che implementano `DataLayer` si sottoscrivono per ricevere aggiornamenti da parte di un singolo `TimeController`, che quindi funge sia da controllore della simulazione sia da *Soggetto*².

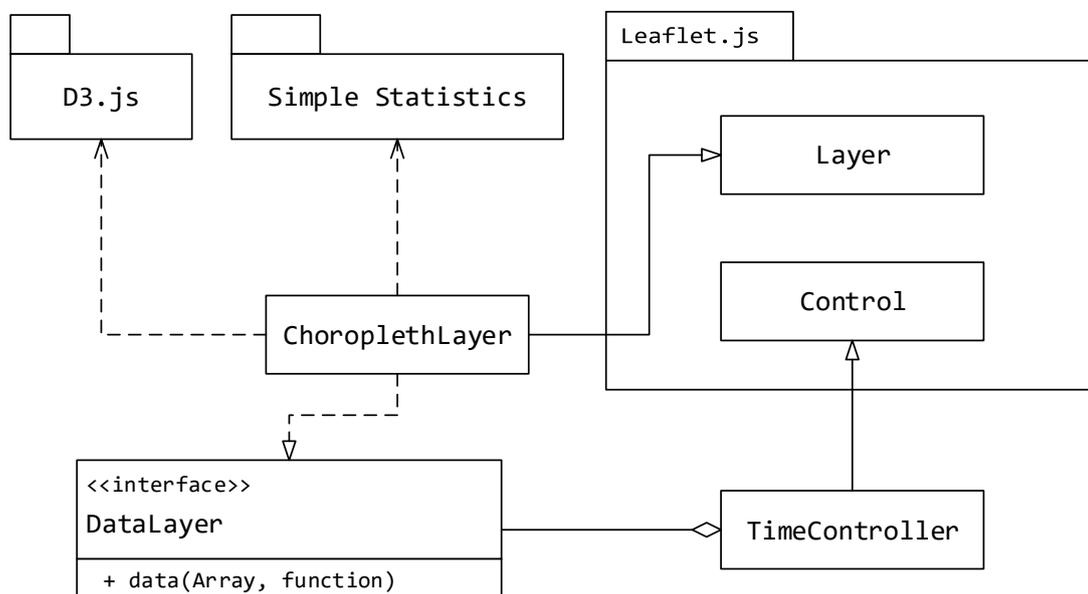


Figura 3.1: Schema UML completo dei componenti del sistema e delle loro dipendenze. Ad uno stesso controllore possono essere associati più livelli in attesa di aggiornamenti. La dipendenza dalla libreria *Simple Statistics* può essere omessa, qualora non si utilizzino le funzionalità descritte nella sezione 3.3.2.

In sintesi, i compiti assegnati ad ogni componente sono:

- **TimeController:** mantiene l'intero set di dati in ingresso al sistema; gestisce la riproduzione della simulazione e il relativo stato corrente, nonché tutte le interazioni dell'utente con l'interfaccia della simulazione; si occupa di mantenere aggiornati i livelli che si sono

²https://en.wikipedia.org/wiki/Observer_pattern

iscritti, inviando loro un set di dati filtrato, contenente solamente dati relativi a eventi che si sono già verificati al tempo corrente della simulazione.

- **ChoroplethLayer**: crea una mappa coropletica in sovrimpressione; per ogni aggiornamento ricevuto, adatta lo stile delle singole aree in base al contenuto del set di dati ricevuto dal controllore; gestisce eventuali interazioni dell'utente con le aree disegnate (ad esempio click, ecc.); offre alcuni metodi statici per la suddivisione dei dati in classi e la creazione di una legenda per l'associazione dei relativi colori.

Ognuno di questi due componenti verrà illustrato con maggiore dettaglio più avanti in questo capitolo.

3.1.2 Uso del sistema

Usati assieme alla libreria di base di Leaflet, questi componenti permettono di creare mappe coropletiche animate ed interattive. L'utente che le visualizza può interagire in molti modi con la presentazione dei dati e il loro stile è altamente personalizzabile, per adattarlo alle diverse esigenze di diversi set di dati. Una mappa interattiva prodotta sfruttando questi componenti è disponibile all'indirizzo <http://wafi.iit.cnr.it/crismap/crismap/mapping>; un esempio di esecuzione è riportato nella Figura 3.2.

3.1.3 Change Blindness e soluzioni adottate

Visti i risultati ottenuti da precedenti studi sull'argomento [13, 16, 17, 18], il sistema che si intende illustrare è stato progettato per garantire la massima interattività possibile della mappa, così da mantenere attiva il più possibile l'attenzione del lettore. Il lettore può andare ad intervenire sulla riproduzione, cambiando la sua velocità, gli istanti di inizio e di fine, mettendola in pausa o specificando manualmente l'istante che intende visualizzare.

È stato scelto invece di non adottare meccanismi di transizione per il passaggio da un set di dati visualizzati ad un altro. Alla base di questa scelta ci sono principalmente due motivazioni.

- La prima è che gli studi precedenti in questo senso hanno ottenuto risultati tra loro contrastanti e non sono stati in grado di stabilire con certezza se l'adozione di tecniche di questo tipo portino effettivi benefici o meno [13].
- La seconda deriva dal fatto che le particolari mappe prodotte con questo sistema permettono di analizzare l'evoluzione di fenomeni in modo pressoché continuo nel tempo: i dati mostrati non passano solamente da uno stato iniziale ad uno finale, bensì vengono mostrati tutti gli eventi che si verificano in ogni dato istante di tempo.

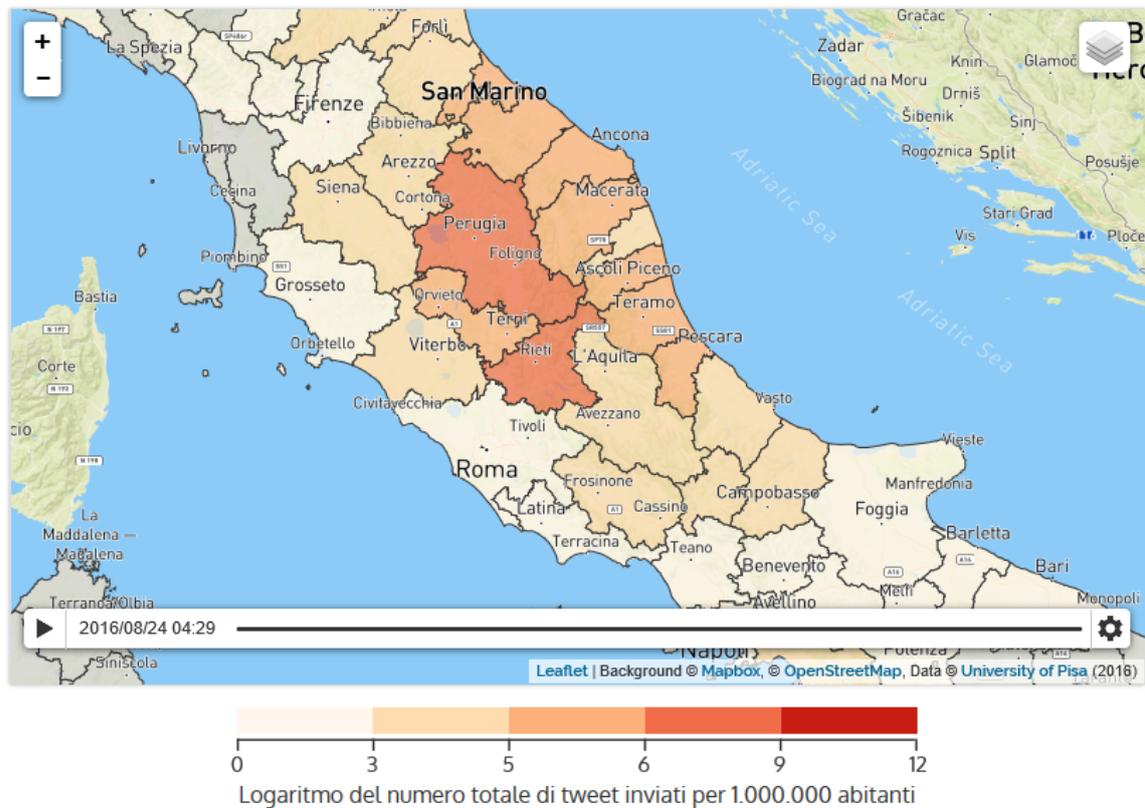


Figura 3.2: *Mappa di crisi relativa al terremoto del Centro Italia del 2016, prodotta a partire dai tweet inviati nella prima ora successiva al sisma, con o senza informazioni relative a entità di danni a cose o persone.*

3.2 Controllore della simulazione

Aggiunto ad una mappa, questo componente permette di creare una simulazione a partire dai dati che gli vengono forniti in ingresso. Questi dati devono essere un insieme strutturato di locazioni geografiche alle quali sono associati eventi localizzabili nel tempo, come ad esempio l'invio di un tweet. Tutti gli eventi vengono inizialmente analizzati per ricavare le informazioni temporali relative, stabilendo i tempi di inizio e di fine della simulazione da realizzare.

Non è richiesta una struttura obbligatoria da usare per il set di dati in ingresso al controllore. Questo poiché oltre al set di dati è richiesto di passare anche delle funzioni di accesso che permettano di ottenere le informazioni di interesse a questo componente.

Il controllore si presenta graficamente come riportato in Figura 3.3.



Figura 3.3: *Forma grafica del controller della simulazione.*

3.2.1 Invio degli aggiornamenti

Quando l'utente avvia la riproduzione, questo componente filtra gli eventi temporali per ogni locazione in base al tempo correntemente visualizzato; il filtraggio prevede di eliminare tutti gli eventi che si sono verificati dopo il tempo corrente della simulazione. Dopo aver ottenuto un set di dati filtrato, il controllore invia ad ogni livello che si è iscritto un aggiornamento contenente il set ottenuto. In questo modo, i singoli livelli possono rappresentare dati in modo incrementale secondo la visualizzazione che ritengono più appropriata.

In verità, per evitare aggiornamenti inutili che appesantirebbero notevolmente il sistema, il controllore invia aggiornamenti ai livelli che ne hanno fatto richiesta solo quando avviene un effettivo mutamento nel set di dati corrente. Quindi, se in un certo intervallo della riproduzione capita che nessuno abbia inviato un tweet, il controllore ritarderà l'invio di aggiornamenti ai vari livelli fino a quando non vi sarà una differenza di almeno un tweet rispetto all'aggiornamento precedente. Questo accorgimento permette di alleggerire il carico del sistema sul browser ospite.

3.2.2 Controllo dell'esecuzione e cambio dei parametri

Il controllore permette all'utente di avviare e fermare la riproduzione, di saltare a un tempo qualsiasi (muovendo lo slider) e cambiarne parametri come i tempi di inizio e di fine della simulazione e la velocità di riproduzione. Per la modifica di questi parametri basta accedere al menù a tendina, riportato in Figura 3.4. Qualora la finestra temporale non venga specificata manualmente, verrà impostata come compresa tra i tempi di invio del primo e dell'ultimo tweet contenuti nel set di dati corrente.

La velocità predefinita di riproduzione di un evento prevede che ogni secondo trascorso nella riproduzione equivalga a un minuto. Potrebbe però essere necessario modificare via software questo parametro per simulazioni di fenomeni che si protraggono a lungo nel tempo (ad esempio per molte ore).

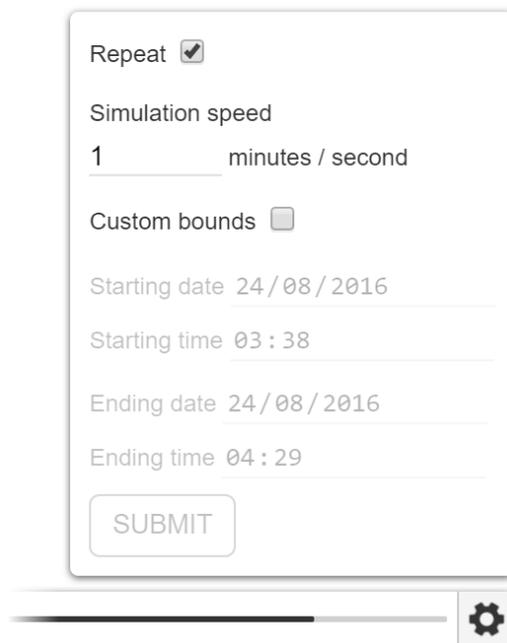


Figura 3.4: Forma grafica del menù di controllo della simulazione. Le opzioni disponibili permettono di rieseguire la simulazione una volta terminata, variare la velocità di esecuzione e la finestra temporale visualizzata.

Tutti i parametri ed i comandi che possono essere impartiti al controllore da un utente possono anche essere eseguiti da chi lo usa per creare una mappa. In questo modo è possibile creare ad esempio visualizzazioni personalizzate, in cui si mette in evidenza l'evoluzione dei dati di una specifica finestra temporale, oppure in cui viene cambiata la velocità di riproduzione predefinita.

3.2.3 Cambio del set di dati

Il set di dati sui quali opera il controllore può essere cambiato durante l'esecuzione della simulazione in corso, senza doverla interrompere, permettendo di confrontare le differenze tra set di dati in ingresso differenti, a parità di posizione geografica e tempo correntemente visualizzato.

Ogni volta che questo set viene modificato, questo componente esegue alcune visite preliminari dei dati ricevuti, calcolando così alcune statistiche temporali di base del set; non appena questa operazione viene completata, esso invia un aggiornamento immediato a tutti i livelli che ne hanno fatto precedentemente richiesta.

È opportuno notare comunque come questa operazione possa essere relativamente onerosa, a seconda della mole dei dati che si intende sostituire.

3.3 Livello coropletico

Questo componente crea un livello che si sovrappone alla mappa e che mostra i dati ricevuti come mappa coropletica. È possibile configurare completamente lo stile degli elementi della mappa e impostare *listener* per reagire alle azioni dell'utente, come il click su di una porzione della mappa, eccetera. Il livello coropletico offre una interfaccia per l'aggiornamento del set di dati correntemente visualizzato compatibile con quella richiesta dal controllore della simulazione. Per disegnare la mappa in sovrapposizione e modificarne in corsa proprietà di stile, questa classe sfrutta la libreria open source *D3.js*³.

Questo componente, grazie a Leaflet, gestisce autonomamente eventi di zoom o movimento della mappa, poiché la porzione della zona visualizzata correntemente si adatta automaticamente agli spostamenti della mappa sottostante, nonché allo scatenarsi di eventi come zoom e similari.

3.3.1 Normalizzazione

Il sistema permette di mostrare a video dati di qualunque natura e per questo non impone limitazioni di sorta alla scala che si vuole impiegare per rappresentarli. Come già abbiamo discusso nel capitolo precedente, è comunque consigliabile adottare un meccanismo di normalizzazione che permetta di confrontare tra loro i dati rappresentati in ogni singola area, a prescindere dalla sua estensione o dalla sua densità di popolazione. L'uso di un meccanismo di questo tipo permette quindi di evidenziare situazioni a rischio anche per comuni scarsamente popolati, dove in genere il numero di tweet raccolti risulta essere relativamente basso [7].

Inoltre, bisogna tenere presente che l'area entro la quale è possibile avvertire, ad esempio, un terremoto è genericamente assai più estesa di quella nella quale possono verificarsi dei danni. Alcuni dei messaggi che verranno raccolti da Twitter potrebbero avere quindi importanza assai minore nella stima dei danni, man mano che la posizione cui si riferiscono si allontana dall'epicentro del fenomeno considerato; questo è dunque un altro punto che potrebbe essere tenuto in considerazione nella normalizzazione delle informazioni.

³<https://d3js.org/>

Poiché il sistema richiede che sia il suo utilizzatore a specificare la tecnica di calcolo impiegata per individuare la classe di appartenenza di una data area, è possibile adottare il sistema di normalizzazione che meglio si adatta ai particolari dati che si intende visualizzare.

3.3.2 Classificazione dei dati

Nella sezione precedente abbiamo già esposto l'importanza della classificazione dei dati per la loro rappresentazione sotto forma di mappa coropletica. Esistono diverse tecniche che permettono di individuare correttamente il numero delle classi in cui suddividere i dati a disposizione e la loro rispettiva estensione. L'obiettivo di questa operazione è quello di minimizzare la varianza "intra-classe" e massimizzare quella "inter-classe". Il sistema implementa tre delle tecniche principali di suddivisione dei dati, quali:

- *Suddivisione per intervalli fissi*: le classi hanno tutte la stessa ampiezza (esempio: 0-10, 10-20, 20-30, e così via); questa tecnica, sebbene facile da realizzare, è valida quando l'insieme dei dati è più o meno uniformemente distribuito tra tutte le classi.
- *Suddivisione dei dati in quantili*: questa tecnica cerca di assegnare ad ogni classe lo stesso numero di porzioni della mappa, arrangiando l'estensione delle classi stesse di conseguenza; se si vuole suddividere un set di 15 aree in 3 classi, allora ogni classe otterrà esattamente 5 aree. Il problema di questa suddivisione è che rischia di produrre range molto disomogenei tra loro (esempio: 1-4, 4-9, 9-250), spesso separando aree con valori tra loro relativamente simili e raggruppando aree con valori assai diversi tra loro, che è il contrario dell'obiettivo che ci siamo proposti. Può essere comunque utile con particolari set di dati.
- *Suddivisione per interruzioni naturali*: prevede di individuare punti di discontinuità nella distribuzione dei dati, in cui porre i limiti delle classi; il risultato finale è una suddivisione in classi con la massima omogeneità interna, che è il nostro obiettivo. Per contro, non funziona molto bene qualora la distribuzione dei dati sia omogenea ed il risultato varia per ogni set di dati che si sceglie in ingresso, rendendo difficile il confronto tra più carte con dati differenti.

Poiché non è possibile stabilire a priori un sistema di classificazione che sia universalmente migliore degli altri per ogni set di dati, si è scelto di rendere disponibili quelli di uso più comune, lasciando all'utilizzatore la scelta del metodo da adottare. Inoltre, è possibile anche non classificare i dati affatto, assegnando ad ogni dato contenuto nel set uno specifico colore, secondo una scala lineare.

Per eseguire alcuni calcoli statistici necessari alla suddivisione dei dati (ad esempio per il calcolo degli intervalli per suddivisioni naturali è richiesto di calcolare varianza, ecc.), il sistema si serve di una libreria di analisi statistica open source chiamata *Simple Statistics*⁴.

Suddivisione in classi e mappe animate

Animare una mappa coropletica richiede di fare alcune considerazioni extra sul numero di classi in cui si intende suddividere il set dei dati da visualizzare. Aumentando il numero delle classi, infatti, aumenta di conseguenza il numero delle tipologie di transizioni differenti che sono possibili in uno stesso istante e di conseguenza anche la frequenza con la quale i dati transizionano tra le varie classi: con 3 classi ad esempio le transizioni possibili sono dalla 1 alla 2, dalla 1 alla 3, dalla 2 alla 3 e le relative transizioni inverse.

Maggiori sono le tipologie di transizioni che il lettore deve cercare di tenere d'occhio, maggiore è la concentrazione che gli è richiesta tenere traccia dei cambiamenti della mappa [13]. Per questo, in letteratura è stato suggerito di limitare in genere il numero di classi impiegate per la realizzazione di una mappa dinamica a due o tre [18], in quanto un numero superiore di classi rischia di limitare la quantità delle informazioni che possono essere correttamente interpretate [13].

Poiché questo suggerimento è basato univocamente sul numero di tipologie di transizioni che possono verificarsi, è possibile comunque adottare un numero superiore di classi, nel caso in cui alcune transizioni da una data classe ad un'altra non possano in alcun modo verificarsi.

Scelta dei colori

Per evitare eventuali problemi di interpretazione della mappa dovuti a cattive percezioni dei colori, il set di colori impiegato per riempire le singole aree della mappa coropletica può essere scelto tra uno qualsiasi di quelli messi a disposizione su ColorBrewer2.org⁵, indipendentemente dal tipo di classificazione (o nessuna classificazione) adottata.

Generazione della legenda

Nel caso in cui si scelga di adottare il meccanismo fornito per la classificazione dei dati e la scelta dei colori, è possibile anche generare in modo automatico una legenda da aggiungere in un luogo a piacere della pagina web ospite della mappa. Questo è fondamentale per una corretta interpretazione della mappa: chi la legge deve essere in grado di comprendere che tipo

⁴<http://simplestatistics.org/>

⁵Vedi l'Appendice A.

di informazione associazione alle varie sfumature di colore; non è possibile dare per scontato che un lettore la interpreti sempre nel modo corretto.

Nella Figura 3.2, in basso, è possibile vedere una legenda costruita secondo il meccanismo offerto.

3.3.3 Zoom e cambio della visualizzazione

È possibile che per mappe caratterizzate da una alta densità di aree da rappresentare possano verificarsi problemi di fluidità del sistema durante la riproduzione di una simulazione. Per esempio, in Italia ci sono 7998 comuni⁶; considerando che per ogni comune potrebbero essere spediti un certo numero di tweet, aggiornare per ogni frame il colore di riempimento o altre caratteristiche delle forme di tutti i comuni italiani potrebbe essere computazionalmente oneroso per un browser.

Per evitare questo, questa classe si preoccupa sia di disegnare sulla mappa solo le aree effettivamente visibili correntemente sullo schermo⁷, sia di cambiare il tipo di mappa da visualizzare a seconda del livello di zoom (per esempio passando da una mappa dei comuni a una delle province o viceversa). Così facendo il problema della fluidità può essere risolto (entro limiti ragionevoli nella mole dei dati) tramite una operazione iniziale di messa a punto di questi parametri, in modo da valutare a quali livelli di zoom assegnare tali cambiamenti di mappa.

Un esempio di questo comportamento è riportato in Figura 3.5.

Ricalcolo del set di dati

Passare da una mappa ad un'altra in seguito ad operazioni di zoom può dover richiedere di ricalcolare il set di dati passato al controllore, in modo che questi possano essere utilizzati correttamente con la nuova visualizzazione. Supponendo, ad esempio, che il set di dati mantenuto dal controllore associ ad ogni comune l'insieme totale dei tweet che lo menzionano, passare ad una visualizzazione per province richiede di avere a disposizione un nuovo set di dati in cui i tweet sono invece raggruppati per provincia.

Questo componente permette di impostare un *listener* che viene invocato ogniqualvolta si verifica una transizione da una visualizzazione ad un'altra, che può essere dunque impiegato per cambiare il set di dati corrente. La pratica più consigliata è quella di calcolare preventi-

⁶Fonte: Istat, Febbraio 2016.

⁷È possibile fare in modo che la porzione disegnata sullo schermo sia più grande della mappa visibile, con una certa percentuale, per permettere uno scorrimento più fluido verso aree limitrofe.

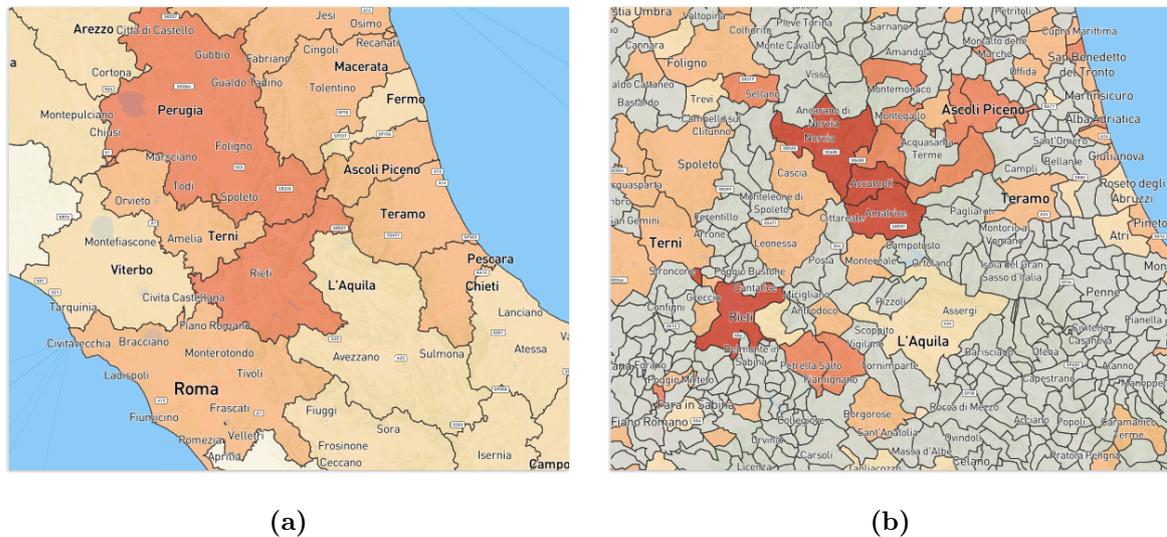


Figura 3.5: *Stessa mappa per diversi livelli di zoom. Nella (a) vengono visualizzati i confini delle province coi relativi dati; aumentando lo zoom di un livello, la mappa passa ai confini dei singoli comuni (b) diventando quindi più dettagliata.*

vamente i set di dati da usare per ognuna delle diverse visualizzazioni possibili (per comune, provincia, regione, ecc.) e di usare questo *listener* per cambiare il set corrente con quello richiesto (già pronto per l'uso).

Nella Figura 3.6 è riportato il diagramma di sequenza che illustra le operazioni appena descritte.

3.4 Prestazioni del sistema

Poiché il sistema deve essere impiegato all'interno di pagine web, è necessario che questo sia sufficientemente leggero da poter essere eseguito da un browser, all'occorrenza anche su di un dispositivo mobile come uno smartphone o un tablet, le cui capacità di calcolo sono relativamente ridotte. Alcune delle tecniche illustrate (per esempio il cambio della mappa con lo zoom, ecc.) sono state adottate per limitare il più possibile il carico applicativo sul browser ospite.

Comunque, per moli di dati relativamente grandi, è possibile che questo non sia sufficiente a mantenere un adeguato livello di fluidità e reattività del sistema. In generale, quindi, è consigliabile eseguire dei test per valutare queste caratteristiche con i particolari dati di interesse. Qualora le prestazioni rilevate si rivelino inferiori agli standard preposti, si consiglia

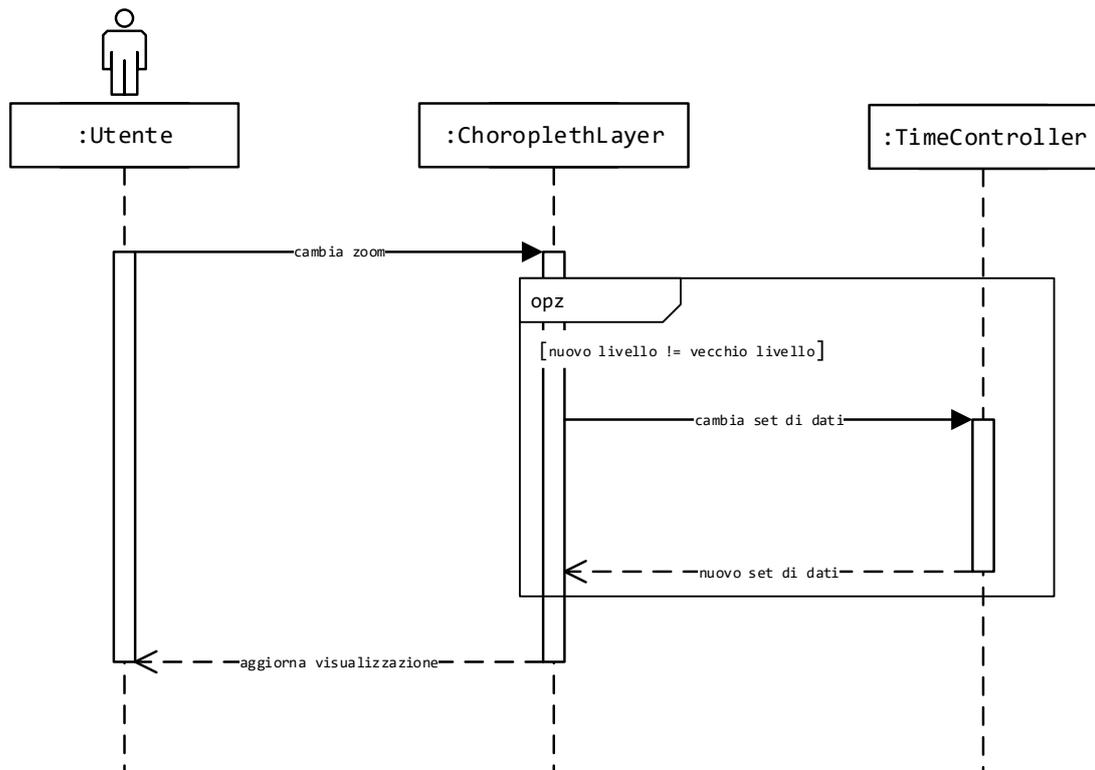


Figura 3.6: Diagramma di sequenza di una operazione di ricalcolo e sostituzione del set di dati corrente.

di adottare un meccanismo di sommarizzazione delle informazioni, in modo da alleggerire il carico dei dati che devono essere elaborati.

4. Caso di studio

In questa sezione viene mostrato come il sistema esposto nella Sezione 3 sia stato impiegato per produrre mappe di crisi relative al caso di studio del terremoto del Centro Italia del 2016¹.

4.1 Il terremoto del Centro Italia

Il fenomeno in questione è iniziato alle 3:36 (ora italiana) del 24 Agosto 2016, quando si è verificata la prima scossa, di magnitudo 6.0 e con epicentro nel comune di Accumoli, in provincia di Rieti. Diverse altre scosse, di gradi anche superiori al quarto, sono state registrate nelle zone limitrofe nel resto della notte e in parte anche nei giorni successivi.

Al momento in cui viene redatto questo documento, i dati relativi all’entità totale dei danni e delle vittime definitivi non sono ancora stati rilasciati. In particolare, non sono ancora disponibili mappe che rappresentino la “verità” da usare per valutare il corretto funzionamento del sistema. Ci si dovrà quindi accontentare di trarre alcune considerazioni preliminari, alcune delle quali di carattere qualitativo, in attesa della pubblicazione di dati certi.

4.2 Fasi del lavoro svolto

4.2.1 Raccolta e geoparsing dei dati

Nell’introduzione di questo documento abbiamo messo in evidenza la necessità di produrre appena possibile mappe di crisi per mezzo di strumenti automatici, in modo da poter organizzare subito i soccorsi in modo adeguato.

Per testare l’affidabilità del sistema in situazioni di criticità come queste si è dunque deciso di creare una mappa di crisi a partire dai tweet relativi al terremoto del Centro Italia che sono stati inviati nella prima ora successiva alla prima scossa. I tweet raccolti sono stati quindi immagazzinati in un database relazionale MySQL e in seguito sottoposti uno per uno

¹https://it.wikipedia.org/wiki/Terremoto_del_Centro_Italia_del_2016

all'analisi di tre annotatori semantici, quali TAGME [21], Dexter [22] e DBPedia Spotlight [23]. Questi strumenti espongono una interfaccia di tipo *RESTful*² che, una volta interrogata, restituisce la risposta all'interno di un oggetto in formato JSON; questo oggetto contiene per ogni entità individuata dall'annotatore il collegamento alla pagina relativa nella base di conoscenza DBPedia e la confidenza di questa associazione. Tra queste entità saranno dunque presenti anche tutte le locazioni geografiche riferite nel testo del messaggio analizzato.

Una volta uniformate queste risposte tra loro è stato necessario eseguire un filtraggio dei risultati ottenuti, mantenendo solamente risultati relativi a locazioni geografiche. Individuate queste entità viene poi interrogata DBPedia per ricavare le loro coordinate geografiche.

Per ogni tweet possono dunque essere annotati più luoghi geografici. Questo pone una ambiguità nella produzione della mappa dei tweet inviati; perciò viene scelto, tra quelli disponibili, quello con distanza minore in linea d'aria dall'epicentro del terremoto. Infine, qualora il luogo ottenuto sia ad esempio una frazione, si è cercato di ottenere il nome del comune italiano contenente tale luogo; qualora nessun comune italiano possa essere associato il luogo indicato, il tweet viene quindi scartato.

I risultati di tutte queste operazioni vengono salvati in un file JSON, in un formato compatibile con quello atteso dal controllore fornito dal sistema illustrato.

4.2.2 Realizzazione della mappa

Una volta ottenuti i dati è stata creata una pagina web per ospitare la mappa risultante. Questa pagina contiene una singola mappa creata con Leaflet, sulla quale possono essere proiettati i dati ottenuti dai vari annotatori. In particolare, i dati ottenuti sono stati suddivisi per annotatore, col fine di confrontare tra loro i rispettivi risultati.

Nella Figura 4.1 è riportato l'insieme di controlli che permette di passare dalla visualizzazione dei dati di un annotatore ad un altro.



Figura 4.1: *Bottoni che permettono di cambiare il set di dati correntemente visualizzato. Ad ogni bottone è associato il risultato prodotto dall'omonimo annotatore.*

La posizione iniziale della mappa è centrata sull'epicentro del terremoto e il livello di zoom iniziale mostra una suddivisione del territorio italiano per province. La mappa sfrutta il mec-

²https://en.wikipedia.org/wiki/Representational_state_transfer

canismo offerto dal livello coropletico per passare da una visualizzazione ad un'altra tramite zoom. Per stabilire il livello di zoom minimo/massimo da associare alle varie visualizzazioni si è cercato di ottenere un giusto compromesso tra il livello di dettaglio delle informazioni deducibili dalla mappa visualizzata (ovvero senza dover intervenire con spostamenti, ecc.) e la reattività/fluidità del sistema durante la riproduzione della simulazione.

Sotto alla mappa è stata inserita, inoltre, una legenda che indica per ogni colore usato la classe associata; in questo caso è stato scelto di adottare una suddivisione in classi per interruzioni naturali. In verità, tutti i metodi di ripartizione per classi resi disponibili dal sistema producono (con questi specifici set di dati) suddivisioni tra loro relativamente simili. La scelta è stata dunque fatta in modo arbitrario.

I dati sono stati normalizzati per numero di abitanti e la scala impiegata è di tipo logaritmico. La scala e la legenda associata sono generate automaticamente dal sistema e variano a seconda del set di dati visualizzato.

4.3 Considerazioni sulle mappe ottenute

Le mappe ottenute sono reperibili all'indirizzo <http://wafi.iit.cnr.it/crismap/crismap/mapping/>. Sebbene non sia possibile (al momento) effettuare confronti con mappe ufficiali relative all'entità di danni a cose o persone, è possibile trarre alcune considerazioni di carattere qualitativo³.

I risultati dei tre annotatori sono tra loro relativamente simili da un punto di vista qualitativo. Tutti e tre evidenziano correttamente come tra le zone più colpite i comuni di Amatrice e Accumoli: alla fine della simulazione, a questi comuni è associata la colorazione più scura della scala. Anche il comune di Norcia è stato classificato tra le zone più danneggiate, coerentemente con i primi risultati dei sopralluoghi, che riportano numerosi danni ad abitazioni e altri edifici, anche se nessun morto.

A molte delle zone limitrofe a quelle appena citate sono state assegnate tonalità vicine a quella più critica; il comune di Arquata del Tronto, anch'esso tra i più colpiti dal sisma, rientra tra queste⁴.

³Le valutazioni dei danni a cose o persone riportate qui sono basate su quanto riportato dalla sezione "Danni e vittime" della pagina https://it.wikipedia.org/wiki/Terremoto_del_Centro_Italia_del_2016, consultata in data 22 Settembre 2016.

⁴Eccezion fatta per Dexter, che non ha riconosciuto il comune affatto. Questo può essere dovuto al fatto che Dexter attinge come base di conoscenza dalla versione in lingua inglese di DBPedia.

5. Conclusioni

In questo documento abbiamo illustrato i principali limiti dei metodi tradizionalmente impiegati per creare mappe di crisi. In particolare, abbiamo notato come la staticità di queste mappe limiti fortemente la gamma di informazioni che può essere ricavata da una loro lettura.

Il sistema presentato in questo lavoro permette di analizzare i dati raccolti dall'attività di Social Sensing in modo da poterne ricavare un numero di informazioni nettamente superiore a quello ottenibile con mappe statiche. Perseguendo il nostro obiettivo di concentrare l'attenzione del lettore sulla mappa nel corso della animazione, il sistema realizzato è stato provvisto di una serie di funzionalità che permette al lettore di controllare completamente l'evolversi del suo contenuto nel tempo e di modificare il livello di dettaglio delle informazioni presentate.

5.1 Sviluppi futuri

Nonostante i risultati ottenuti, il sistema può essere migliorato in vari modi:

- La modularità con la quale è stato progettato permette di riutilizzare il controllore della simulazione per realizzare visualizzazioni di tipo differente da quella della mappa coropletica; per farlo è sufficiente che il nuovo livello realizzato implementi l'interfaccia descritta nella sezione 3.1.1 e si sottoscriva per la ricezione degli aggiornamenti da parte del controllore.
- L'alto grado di personalizzazione lasciato all'utilizzatore del sistema permette di aggiungere funzionalità mediante la definizione di nuove reazioni a interazioni dell'utente con la mappa. Ad esempio, il click su di una porzione di territorio può scatenare la creazione di una visualizzazione accessoria che permette di ottenere informazioni aggiuntive relative alla zona selezionata, come il contenuto dei tweet ad essa associati, ecc.

Inoltre, le funzionalità del sistema possono essere amplificate introducendo dei meccanismi per la raccolta dei dati in tempo reale, permettendo così di monitorare eventi live e ricevere aggiornamenti automatici sull'evoluzione di un fenomeno attualmente in corso.

Perché questa funzionalità possa essere aggiunta è necessario rivedere un poco il sistema impiegato nel Capitolo 4 nella sua interezza, tenendo presente che la parte descritta qui rappresenta solo l'ultimo passaggio di una sequenza di operazioni necessarie per ogni tweet che viene spedito.

Raccolta del messaggio

Il messaggio deve essere ottenuto dalla base dati di Twitter. Per questa operazione può essere adattato sistema impiegato nella sezione 4.2.1, in modo che scarichi in tempo reale ogni tweet relativo all'evento che si sta monitorando.

Analisi semantica e geotagging

Il tweet deve essere taggato e georiferito con un annotatore semantico. Anche per questa operazione potrebbe essere impiegato quanto già visto nella sezione 4.2.1. Bisogna comunque notare che l'invio di richieste alle versioni online dei vari annotatori richiede del tempo, che potrebbe essere eccessivo per realizzare un meccanismo per la ricezione di aggiornamenti live.

Potrebbe dunque essere necessario un certo investimento in hardware dedicato all'annotazione dei tweet, per evitare ritardi in questa fase del lavoro. Installare in locale vari annotatori semantici permette di avere un maggiore controllo sulle risorse allocate a questi servizi e di evitare di dover fare richieste a server remoti, usati anche da altre applicazioni.

Trasmissione della nuova informazione al client

Il server che si occupa di queste operazioni deve quindi notificare a tutti i client interessati l'arrivo di un nuovo tweet pronto per essere mostrato sulla mappa. Questo può essere realizzato per mezzo di varie tecniche; le due alternative principali sono:

- L'apertura di un socket persistente tra il server e i vari client. In particolare i client web possono utilizzare l'interfaccia offerta dai *WebSocket*¹; ovviamente, client di altre tipologie (app, ecc.) possono sempre aprire un socket tradizionale per comunicare col server.
- Usando la tecnologia di *HTTP Live Streaming* (HLS)² per ricevere aggiornamenti sotto forma di file scaricati via HTTP. Poiché in questa particolare applicazione non vengono trasferite informazioni di qualche rilevanza dal client verso il server, questa soluzione può

¹<https://en.wikipedia.org/wiki/WebSocket>

²https://en.wikipedia.org/wiki/HTTP_Live_Streaming

essere preferibile all'uso di WebSocket, in quanto non è veramente necessario stabilire una connessione full-duplex tra le due parti. Adottare questo meccanismo permette, comunque, di realizzare client anche per dispositivi mobili, poiché, ad esempio, iOS e Android supportano nativamente la ricezione di aggiornamenti tramite HLS.

- Usando i *Server-Sent Events* (SSE)³; questa tecnologia è simile a quella dell'HLS, poiché gli aggiornamenti vengono ricevuti tramite il protocollo HTTP e la comunicazione è possibile nella sola direzione dal server verso i client. La tecnologia SSE è implementata nativamente nella maggioranza dei browser moderni e, qualora non sia nativamente supportata, esistono molte librerie che ne permettono l'utilizzo.

Nel caso in cui si voglia realizzare questa funzionalità, il sistema illustrato in questo lavoro necessita, dunque, di essere adattato perché sia possibile non solo sostituire l'intero set di dati visualizzato "on the fly", ma anche aggiungere nuove informazioni al set corrente in modo incrementale.

³https://en.wikipedia.org/wiki/Server-sent_events

Appendice A. Percezione dei Colori

Poiché nelle mappe coropletiche l'informazione è trasmessa al lettore attraverso intensità e sfumature di colori, risulta necessario, oltre che scegliere accuratamente la suddivisione dei dati nelle classi opportune, associare anche correttamente un colore specifico ad ogni classe, in modo che il risultato sia il più chiaro possibile.

Un problema che si verifica abbastanza frequentemente nella lettura di mappe coropletiche è quello del *contrasto simultaneo*, ovvero l'aumento di luminosità dovuto all'accostamento di due colori complementari⁴. In questo caso, alcune possono sembrare colorate di una sfumatura più intensa di quella effettiva, a causa della colorazione più scura delle aree loro confinanti, e viceversa. Il set di colori impiegato per creare mappe coropletiche deve essere tale da eliminare il più possibile problemi di questo tipo.

Fortunatamente, il tema della scelta del set di colori per mappe tematiche e coropletiche è un argomento abbastanza noto in letteratura; in particolare, i lavori di Brewer [24, 25] in questo senso hanno portato alla realizzazione di uno strumento che da ormai alcuni anni è diventato quasi standard nella produzione di mappe di questo tipo, ovvero ColorBrewer [26]. Questa web app, disponibile all'indirizzo <http://colorbrewer2.org>, permette di confrontare diversi set di colori dalle caratteristiche note e di filtrare questi set in base alle proprie esigenze e al numero di classi che si vuole adottare; in generale, per ogni set è associato una serie di proprietà, quali:

- Adatto per l'interpretazione dei dati anche da parte di daltonici;
- Adatto ad essere stampato;
- Adatto per essere fotocopiato;
- Adatto ad essere visualizzato su schermi LCD.

⁴https://en.wikipedia.org/wiki/Contrast_effect

Inoltre, lo strumento permette di valutare l'effetto complessivo della scala adottata in modo grafico, valutando eventuali effetti di contrasto simultaneo, anche se questi set di colori dovrebbero essere studiati appositamente per evitare problemi di questo tipo.

Bibliografia

- [1] B. Birregah, T. Top, C. Perez, E. Châtelet, N. Matta, M. Lemercier, and H. Snoussi, “Multi-layer crisis mapping: a social media-based approach,” in *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2012 IEEE 21st International Workshop on*. IEEE, 2012, pp. 379–384.
- [2] M. Avvenuti, S. Cresci, A. Marchetti, C. Meletti, and M. Tesconi, “EARS (Earthquake Alert and Report System): A real time decision support system for earthquake crisis management,” in *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2014, pp. 1749–1758.
- [3] M. Avvenuti, S. Cresci, F. Del Vigna, and M. Tesconi, “Impromptu crisis mapping to prioritize emergency response,” *Computer*, vol. 49, no. 5, pp. 28–37, 2016.
- [4] J. Ziemke, “Crisis mapping: The construction of a new interdisciplinary field?” *Journal of Map & Geography Libraries*, vol. 8, no. 2, pp. 101–117, 2012.
- [5] P. Meier, “Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time,” *Journal of Map & Geography Libraries*, vol. 8, no. 2, pp. 89–100, 2012.
- [6] P. S. Earle, D. C. Bowden, and M. Guy, “Twitter earthquake detection: earthquake monitoring in a social world,” *Annals of Geophysics*, vol. 54, no. 6, 2012.
- [7] S. Cresci, A. Cimino, F. Dell’Orletta, and M. Tesconi, “Crisis mapping during natural disasters via text analysis of social media messages,” in *International Conference on Web Information Systems Engineering–WISE 2015*. Springer, 2015, pp. 250–258.
- [8] A. L. Hughes and L. Palen, “Twitter adoption and use in mass convergence and emergency events,” *International Journal of Emergency Management*, vol. 6, no. 3-4, pp. 248–260, 2009.

- [9] S. E. Middleton, L. Middleton, and S. Modafferi, "Real-time crisis mapping of natural disasters using social media," *IEEE intelligent systems*, vol. 29, no. 2, pp. 9–17, 2014.
- [10] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010, pp. 759–768.
- [11] M. Harrower, "The cognitive limits of animated maps," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 42, no. 4, pp. 349–357, 2007.
- [12] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. Howard, *Thematic cartography and geovisualization*. Pearson Prentice Hall Upper Saddle River, NJ, 2009.
- [13] K. Goldsberry and S. Battersby, "Issues of change detection in animated choropleth maps," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 44, no. 3, pp. 201–215, 2009.
- [14] G. Andrienko, N. Andrienko, and A. Savinov, "Choropleth maps: Classification revisited," in *Proceedings ICA*. Citeseer, 2001, pp. 1109–1219.
- [15] W. R. Tobler, "Choropleth maps without class intervals?" *Geographical analysis*, vol. 5, no. 3, pp. 262–265, 1973.
- [16] D. J. Simons and R. A. Rensink, "Change blindness: Past, present, and future," *Trends in cognitive sciences*, vol. 9, no. 1, pp. 16–20, 2005.
- [17] B. Tversky, J. B. Morrison, and M. Betrancourt, "Animation: can it facilitate?" *International journal of human-computer studies*, vol. 57, no. 4, pp. 247–262, 2002.
- [18] M. Harrower, "Tips for designing effective animated maps," *Cartographic Perspectives*, no. 44, pp. 63–65, 2003.
- [19] S. I. Fabrikant, S. Rebich-Hespanha, N. Andrienko, G. Andrienko, and D. R. Montello, "Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes," *The Cartographic Journal*, vol. 45, no. 3, pp. 201–215, 2008.
- [20] D. J. Simons, S. L. Franconeri, and R. L. Reimer, "Change blindness in the absence of a visual disruption," *Perception*, vol. 29, pp. 1143–1154, 2000.

- [21] P. Ferragina and U. Scaiella, “Tagme: on-the-fly annotation of short text fragments (by wikipedia entities),” in *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010, pp. 1625–1628.
- [22] S. Trani, D. Ceccarelli, C. Lucchese, S. Orlando, and R. Perego, “Dexter 2.0 - an open source tool for semantically enriching data,” in *Proceedings of the 13th International Semantic Web Conference*, ser. CIKM '10, 2014.
- [23] J. Daiber, M. Jakob, C. Hokamp, and P. N. Mendes, “Improving efficiency and accuracy in multilingual entity extraction,” in *Proceedings of the 9th International Conference on Semantic Systems*. ACM, 2013, pp. 121–124.
- [24] C. A. Brewer, “Color use guidelines for mapping,” *Visualization in modern cartography*, pp. 123–148, 1994.
- [25] C. A. Brewer, A. M. MacEachren, L. W. Pickle, and D. Herrmann, “Mapping mortality: Evaluating color schemes for choropleth maps,” *Annals of the Association of American Geographers*, vol. 87, no. 3, pp. 411–438, 1997.
- [26] M. Harrower and C. A. Brewer, “Colorbrewer. org: An online tool for selecting colour schemes for maps,” *The Cartographic Journal*, vol. 40, no. 1, pp. 27–37, 2003.

Ringraziamenti

Ringrazio tutti coloro che mi hanno sostenuto in questi anni e nella realizzazione di questo progetto. Il mio primo e più grande ringraziamento va alla mia famiglia, che da sempre mi sprona a dare il meglio di me stesso, a impegnarmi con costanza e che mi è sempre vicina nei momenti più difficili.

Ringrazio in particolar modo i miei compagni di corso, che mi hanno sopportato e supportato, condividendo con me non solo una serie di lezioni, ma anche risate, cene e altri bei momenti vissuti in questi anni.

Ringrazio Gabriele Baris, per tutta la pazienza che ha dimostrato nelle innumerevoli volte in cui gli ho chiesto consiglio, soprattutto in questi ultimi mesi.

Ringrazio i professori Marco Avvenuti e Maurizio Tesconi, che mi hanno proposto questo lavoro e messo a disposizione le risorse per realizzarlo. Ringrazio gli ingegneri Fabio Del Vigna e Stefano Cresci per tutto il tempo che mi hanno dedicato e per i loro preziosissimi aiuti e consigli.

Per ultimo, ma tutt'altro che per importanza, vorrei dedicare un ringraziamento particolare al mio "compagno d'armi" Gabriele Serra, che non solo ha accettato di gettarsi con me in questa impresa, ma che col suo modo di fare, i suoi consigli e i suoi incoraggiamenti mi ricorda sempre che esiste qualcosa oltre al lavoro a cui dovrei dare uguale importanza.