

**Analisi lessico testuale delle
ordinanze del Commissario Delegato
alla ricostruzione in Emilia-Romagna:
un contributo alla legge nazionale su
emergenze e ricostruzione**

Pasquale Pavone, Margherita Russo

CAPPaper n. 128
novembre 2015



Università di Modena e Reggio
Emilia Facoltà di Economia
Marco Biagi



Università di Bologna
Dipartimento di Scienze
Economiche

CAPP - Centro di Analisi delle Politiche Pubbliche
Dipartimento di Economia Politica - Università di Modena e Reggio Emilia
Ufficio 54 - Ala Ovest
Viale Berengario, 51 41100 Modena - ITALY
phone: +39 059 2056854 fax: +39 059 2056947
email capp@unimo.it

CAPP & DEMB-WP

**Analisi lessico testuale delle ordinanze
del Commissario Delegato alla ricostruzione in Emilia-Romagna:
Un contributo alla legge nazionale su emergenza e ricostruzione**

Lexical-textual analysis of the ordinances
of the Deputy Commissioner for reconstruction in Emilia-Romagna:
A contribution to the national law of emergency and reconstruction

Pasquale Pavone* e Margherita Russo**

Novembre 2015

*Progetto Energie Sisma Emilia

**University of Modena and Reggio Emilia
CAPP, Centre for the Analysis of Public Policies
Address: Viale Berengario 51, 41121 Modena, Italy
email: margherita.russo@unimore.it

Abstract

Il Commissario delegato all'emergenza e alla ricostruzione dopo il sisma in Emilia-Romagna del 20 e 29 maggio 2012, ha gestito gli interventi a favore delle popolazioni colpite attraverso l'emanazione di ordinanze commissariali. In particolare dal 8 giugno 2012 al 23 febbraio 2015, sono state emanate 349 ordinanze. Si è trattato di una produzione normativa, indispensabile per colmare un vuoto legislativo, che ha consentito di superare le incertezze della difficile fase di ripresa dopo lo shock sismico. È opinione diffusa tra gli esperti che un gran numero di quelle ordinanze sono state necessarie per l'assenza di una normativa nazionale che disciplini gli interventi urgenti in caso di calamità naturali. Sulla spinta delle azioni messe in atto in Emilia-Romagna, è stato ripreso il dibattito parlamentare su una legge nazionale sull'emergenza, la cui discussione è attualmente in corso alla Camera dei Deputati.

In questo documento proponiamo un contributo a quel disegno di legge attraverso un'analisi automatica dei testi delle ordinanze in grado di fornire una lettura temporale delle tematiche affrontate e degli atti normativi in esse citati. L'analisi lessico-testuale, effettuata utilizzando il software Taltac2 sul Corpus costituito dalle Ordinanze, mira ad una classificazione automatica delle tematiche e ad un'analisi temporale dei temi affrontati dall'azione normativa nel corso del tempo. Sono state individuate quattro tematiche principali: criteri di concessione dei contributi; gestione e assegnazione delle risorse; opere urgenti per scuole, municipi e chiese; interventi di assistenza alla popolazione. Avendo associato a ciascuna ordinanza a uno dei temi individuati dall'analisi cluster, abbiamo analizzato la sequenza di azioni intraprese nel corso della ricostruzione. I termini che caratterizzano ciascun cluster sono stati utilizzati per ottenere una definizione di classi disgiunte attraverso una categorizzazione fuzzy plurima.

Abstract

In the three years after the 2012 earthquake in Emilia-Romagna, through the enactment of more than 350 ordinances, the Commissioner has structured interventions to cope with emergency and reconstruction. The intense law-making, essential to fill a legal vacuum, has enabled to overcome the uncertainties of the difficult phase of recovery. There is agreement among experts that a large number of those ordinances was due to the absence of national rules governing the urgent intervention in case of natural disasters. On the push of actions taken in Emilia-Romagna, the Italian Parliament has reopened the debate on a national law on emergency after natural disasters.

Through a systematic content analysis of the corpus of ordinances issued in Emilia-Romagna, in this paper we propose a contribution in drafting a law on emergency. Two main strands of analysis have been developed.

In the first one, an automatic text analysis, supported by Taltac2, has provided inputs for a factor analysis and a cluster analysis of the thematic areas covered by the ordinances. Four main topics have been singled out: grant criteria and contributions; management of allocation of resources; urgent works for municipalities, schools and churches buildings; interventions to support population. Having associated each ordinance to one of the four topics, a temporal analysis of the issues addressed during the emergency and reconstruction phase highlights the sequence of actions that were undertaken in Emilia. In a second step, the set of terms characterizing each cluster it is used to obtain a redefinition of disjunctive classification towards a fuzzy multi-class.

Key words: automatic classification; legislative corpus; clustering and network analysis; Law Making; Data Collection, Data analysis, computer software.

JEL: K4 Legal Procedure, the Legal System, and Illegal Behavior; K39 Other; C81 Methodology for Collecting, Estimating, and Organizing Microeconomic Data; Data Access; C88 Other Computer Software; R58 Regional Development Planning and Policy;

Credits

Questo lavoro è stato prodotto nell'ambito del progetto di ricerca applicata (2014-2016) "Energie Sisma Emilia" (www.energie.unimore.it) dell'Università di Modena e Reggio Emilia, cofinanziato da Fondazione Cassa di Risparmio di Modena, Azienda Unità Sanitaria Locale di Modena e Università di Modena e Reggio Emilia.

Si ringrazia il gruppo di ricerca Energia Sisma Emilia dell'Università di Modena e Reggio Emilia (www.energie.unimore.it) per le discussioni e commenti sul progetto di analisi delle ordinanze. Un particolare ringraziamento: a Manuel Reverberi e Marco Ranuzzini per la collaborazione nella creazione del database delle ordinanze e per la disambiguazione degli atti normativi; a Simone Righi per le indicazioni sull'impostazione del data base che sarà analizzato con infomap; a Simone Scagliarini per i commenti sul progetto e per le revisioni apportate al testo con riferimento ai temi normativi; a Monica Palmirani per aver ampliato gli orizzonti di questa ricerca, intrecciando metodi e strumenti per irrobustire le domande e i risultati della ricerca. Ogni responsabilità su quanto riportato nel testo resta da attribuire agli autori.

Il progetto, condotto dai due autori in stretta collaborazione, è stato sviluppato da Pasquale Pavone.

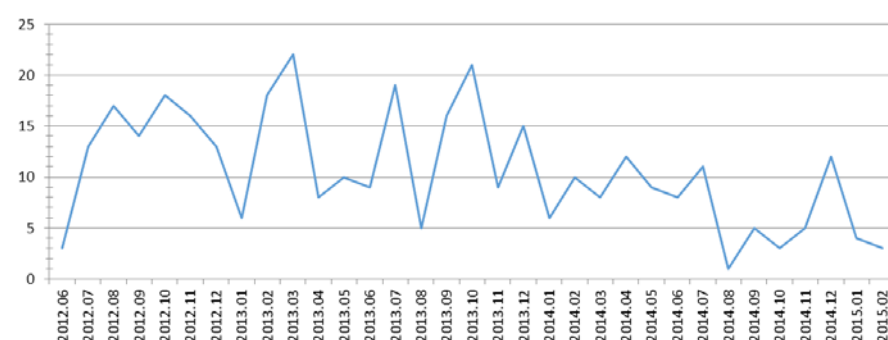
1 Introduzione

1.1 Le ordinanze del Commissario Delegato all'emergenza e alla ricostruzione

Dopo il sisma del 20 e 29 maggio 2012 in Emilia-Romagna, Veneto e Lombardia, il Governo ha emanato d'urgenza il Decreto n.74 del 6 giugno 2012, attribuendo ai presidenti delle regioni colpite dal sisma i poteri di definire gli interventi necessari per l'emergenza e la ricostruzione¹.

In Emilia-Romagna, il Commissario delegato ha gestito gli interventi a favore delle popolazioni colpite attraverso l'emanazione di ordinanze commissariali, una produzione normativa indispensabile per colmare un vuoto legislativo, che ha consentito di superare le incertezze della difficile fase di ripresa dopo lo shock sismico. Se consideriamo il periodo giugno 2012-febbraio 2015, oggetto di analisi in questo lavoro, sono state emesse 349 ordinanze, in media 10,6 ordinanze al mese, con un minimo di un'ordinanza nell'agosto del 2014 a un massimo di 22 ordinanze nel marzo 2013². Dalla Figura 1 risulta che, a meno del rallentamento in dicembre 2012 e gennaio 2013 ci sono picchi che si succedono a mesi con una minore emissione di ordinanze.

Figura 1 - Numero di ordinanze per mese (giugno 2012 - febbraio 2015)



Fonte: nostra elaborazione sul corpus di testi delle ordinanze del Commissario Delegato Emilia-Romagna

È opinione diffusa tra gli esperti che un gran numero di quelle ordinanze sono state necessarie per l'assenza di una normativa nazionale che disciplini gli interventi urgenti in caso di calamità naturali. Sulla spinta delle azioni messe in atto in Emilia-Romagna, è stato ripreso il dibattito parlamentare su una legge naziona-

¹ La Legge 24 febbraio 1992, n. 225, attribuiva al capo di Protezione Civile il potere di emanare "ordinanze in deroga ad ogni disposizione vigente, nei limiti e secondo i criteri indicati nel decreto di dichiarazione dello stato di emergenza e nel rispetto dei principi generali dell'ordinamento giuridico".

² Alla data di pubblicazione di questo lavoro, il 20 novembre 2015, le ordinanze emesse sono 391.

le sull'emergenza, la cui discussione è attualmente in corso alla Camera dei Deputati.

Nell'ambito del progetto di ricerca Energia Sisma Emilia l'analisi degli atti normativi è stata avviata per acquisire la base informativa essenziale per analizzare il contesto normativo del processo di ricostruzione, e per studiare il processo organizzativo messo in atto per far fronte all'emergenza. Da una prima analisi temporale della produzione normativa (rappresentata nella Figura 1) è scaturito l'interesse ad approfondire le caratteristiche tematiche delle ordinanze per tracciare gli ambiti di intervento su cui si sviluppa l'azione nella fase di emergenza e di ricostruzione.

Costruito il database delle ordinanze³, sono stati avviati vari approfondimenti complementari per analizzare la produzione normativa messa in atto in Emilia-Romagna. Un primo approfondimento, presentato in questo lavoro, propone una classificazione dei temi contenuti nelle ordinanze, utilizzando l'analisi automatica dei testi, che ci consente anche di individuare gli atti normativi citati nelle ordinanze del Commissario delegato. Un secondo approfondimento riguarda la rete di norme citate e l'analisi delle tipologie di modifiche intervenute nel corso del tempo. All'analisi delle reti di atti normativi sono dedicati due contributi sviluppati in collaborazione con Simone Righi e con Monica Palmirani. Nell'analisi delle modifiche degli atti normativi, svolta con tecniche di informatica giuridica da Monica Palmirani, Ilaria Bianchi, Luca Cervone e Francesco Draicchio, del CIRSIFID dell'Università di Bologna, si utilizza il vocabolario di categorie e sottocategorie tematiche emerse nell'analisi automatica dei testi presentata in questo lavoro. Nel complesso gli approfondimenti in corso offrono un contributo originale al disegno di legge nazionale sull'emergenza e la ricostruzione.

1.1 La struttura di questo lavoro

Gli atti emanati dal Commissario Delegato considerati nella nostra analisi coprono un arco temporale che va dal giugno del 2012 al febbraio del 2015. All'interno di essi si possono riscontrare diverse tipologie di atti, in particolare:

- Ordinanze: atti normativi. Numerose Ordinanze sono modifiche o integrazioni di atti precedenti. In alcuni casi, su materie particolarmente rilevanti (ad esempio contributi alla ricostruzione delle imprese o delle abitazioni

³ I testi ufficiali delle ordinanze commissariali - aventi valore legale - sono pubblicati sul Bollettino ufficiale della Regione Emilia-Romagna. Nella sezione del sito della Regione Emilia-Romagna "Terremoto, la ricostruzione" <http://www.regione.emilia-romagna.it/terremoto/gli-atti-per-la-ricostruzione> sono scaricabili i file pdf delle ordinanze degli anni 2014 e 2015. Al momento in cui abbiamo creato il nostro database erano invece accessibili da quella sezione anche i file delle ordinanze del 2012 e del 2013. I documenti presenti nella sezione del sito hanno scopo meramente informativo e divulgativo. I testi ufficiali delle ordinanze commissariali - aventi valore legale - sono pubblicati sul Bollettino ufficiale della Regione Emilia-Romagna.

private) sono stati emanati dei Testi Coordinati di Ordinanze, atti normativi che raccolgono in un unico testo il provvedimento originario comprese le modifiche.

- Decreti e Circolari: atti amministrativi che hanno lo scopo di chiarire e spiegare le disposizioni contenute nelle Ordinanze o in altre norme. Essi sono soprattutto ad “uso interno” per gli addetti ai lavori.

Per lo svolgimento dell'analisi si è scelto di considerare come unità testuali esclusivamente le ordinanze emanate dai Commissari delegati.

Dopo aver richiamato i principali elementi che caratterizzano l'analisi automatica dei testi (sezione 2) e il corpus di testi oggetto di analisi (sezione 3), i principali elementi di trattamento del corpus sono descritti nella sezione 4. Hanno riguardato l'esplorazione del corpus complessivo tramite l'estrazione del linguaggio peculiare, dei verbi peculiari, l'individuazione di multiword, il conteggio delle citazione dei Comuni richiamati nelle ordinanze (al fine di individuare l'intensità normativa che li riguarda). L'individuazione degli atti normativi citati in tutto il corpus (sezione 5) si è basata su una serie di query che consentirà al gruppo di ricerca di aggiornare l'analisi includendo le ordinanze emesse successivamente al periodo oggetto di analisi. L'identificazione di tali atti normativi ci ha consentito la costruzione della matrice ordinanze/citazioni che verrà elaborata con Infomap in altra parte della ricerca⁴. Attraverso una clusterizzazione non supervisionata dei documenti, nella sezione 6 si è proceduto all'individuazione delle tematiche generali presenti nelle ordinanze e, a partire dalla classificazione disgiunta ottenuta con la Cluster Analysis, è stata effettuata la categorizzazione fuzzy delle ordinanze e sono messi in evidenti i principali spunti emerse dall'analisi con riferimento alle peculiarità temporale della produzione normativa nei diversi ambiti tematici individuati. La sezione 7 conclude e presenta le principali linee di sviluppo della ricerca in corso.

2. L'analisi automatica dei testi

In una prospettiva di analisi qualitativa e quantitativa dei contenuti, proprietà e caratteristiche dei testi, le Tecniche Automatiche di Analisi del Testo permettono di non leggere il testo, ma di interpretare le rappresentazioni delle informazioni contenute in esso. Tali tecniche rappresentano il contesto naturale per l'applicazione degli strumenti di TextMining (Sullivan, 2001) (Bolasco, et al., 2005) (Zanasi, 2005).

Il Text Mining è un'area di ricerca multidisciplinare che, combinando con uguale importanza strumenti della Linguistica Computazionale, dell'Information

⁴ Questa analisi, a cui lavorano Simone Righi, Pasquale Pavone e Margherita Russo, utilizzerà anche la classificazione delle modifiche elaborata da Palmirani et al. (2015).

Retrieval e della Statistica, mira all'estrazione dell'informazione d'interesse presente nei testi scritti in linguaggio naturale.

L'utilizzo di tecniche automatiche di analisi del testo pertanto si rende necessario nei casi in cui la quantità di informazioni disponibili è tale da non permettere la risoluzione manuale dei problemi di classificazione e clustering dei dati (Bolasco, 2013). I *Corpus* di documenti disponibili per un'analisi del contenuto sono sempre di maggiori dimensioni e spesso uno dei problemi essenziali è la corretta attribuzione di una o più categorie tematiche in grado di individuare il contenuto dei diversi testi in analisi.

Nel trattamento automatico di tipo lessicale (TAL) l'oggetto di studio è il lessico, inteso come la terminologia di un corpus di testi. L'analisi del lessico avviene attraverso l'esplorazione delle informazioni strutturate nel database Vocabolario, dove per ogni parola diversa (type) possono essere associate delle annotazioni di vario tipo: linguistico, semantico, quantitativo, statistico. Tali annotazioni sono il risultato del trattamento automatico a diversi step dell'analisi: normalizzazione del testo; tagging grammaticale e/o semantico, calcolo delle sub-occorrenze, calcolo di indici statistici per la selezione del linguaggio. Attraverso l'utilizzo di indici statistici è possibile infatti selezionare determinate parti del vocabolario quali: linguaggio peculiare (sovra/sotto utilizzato rispetto all'uso medio di un lessico di riferimento), linguaggio rilevante (estratto tramite l'indice TFIDF (Salton, 1989) discriminante rispetto ai documenti), linguaggio specifico (caratteristico di sub-testi o partizioni del corpus, intese come modalità di una variabile categoriale). Ciascuna di queste proprietà costituisce un esempio di meta-informazione attribuita al type, recuperabile attraverso l'interrogazione dei corrispondenti campi del database Vocabolario in cui tale informazione è depositata. L'estrazione/selezione delle parti del vocabolario serviranno per “raccontare” le caratteristiche lessicali del corpus evidenziando gli elementi significativi di ciascuna “parte del discorso”, o per “illustrare” nelle mappe fattoriali determinati sotto-insiemi di types.

Nel trattamento automatico di tipo testuale (TAT), l'oggetto di studio è il Corpus inteso come susseguirsi di parole all'interno di una collezione di testi da analizzare, confrontare e categorizzare. L'unità d'analisi è *l'unità di contesto*, ovvero un frammento di testo, sia esso una frase o un intero documento. In analogia con il TAL, ogni unità di contesto costituisce un'entrata nel database Frammenti a cui sono associate sia le modalità delle variabili codificate a priori sia le annotazioni testuali come risultato del TAT. Tali annotazioni possono essere anche in questo caso di vario tipo: linguistico, come l'individuazione di strutture sintattiche o di gruppi ad elementi variabili; semantico, come le categorizzazioni automatiche sulla base di lessici specifici; quantitativo, come il conteggio di occorrenze di una determinata entità o come il calcolo del TFIDF rispetto ad una query. Ad eccezio-

ne delle annotazioni effettuate a partire dal TFIDF, generalmente le annotazioni nel database Frammenti delle unità di contesto avvengono mediante la definizione di query testuali consistenti in *espressioni regolari*⁵(ER). Attraverso le ER infatti si ricercano delle stringhe di testo che di volta in volta possono essere le occorrenze sia di singoli type o entità complesse, sia di loro classi che di relazioni fra classi o fra singoli type e classi. Il risultato di tale elaborazione consiste nel recuperare i frammenti che verificano la query testuale; inventariare la lista delle stringhe estratte; annotare eventualmente i frammenti.

Attraverso l'integrazione nel trattamento automatico dei testi fra il livello di analisi lessicale e quello testuale è possibile definire un sistema ibrido di analisi che impiega in maniera congiunta e reiterata dei dizionari e delle regole, per la ricerca di entità complesse con struttura ad elementi variabili (Bolasco & Pavone, 2010). Nell'ambito dell'estrazione delle informazioni, una struttura ad elementi variabili equivale ad un'entità di tipo concettuale articolata in varie componenti da ricercare ed estrarre da un testo come nel caso esposto nel presente lavoro, delle citazioni degli atti normativi in un testo giuridico.

L'analisi lessico testuale esposta è stata effettuata utilizzando il software Taltac2⁶. La filiera di lavoro è caratterizzata dalle seguenti cinque fasi:

- Raccolta delle Informazioni, consistente nel reperimento delle fonti testuali digitalizzate (ordinanze) dall'apposita sezione del sito della Regione Emilia Romagna⁷;
- Pre-Trattamento dei testi, consistente nella loro tokenizzazione (acquisizione del testo mediante indicizzazione numerica delle parole), per il riconoscimento delle unità d'analisi, e nella costituzione del document warehouse;

⁵ Notazione algebrica che permette di definire in maniera formale e rigorosa dei modelli di stringhe (Lenci, et al., 2005).

⁶ TALTAC sta per Trattamento Automatico Lessico-Testuale per l'Analisi del Contenuto (Bolasco, 2010). Sviluppato fin dal 1999 presso il Dipartimento MEMOTEF dell'Università "La Sapienza" di Roma, è una libreria di programmi che consente il trattamento e l'analisi di un insieme di dati espressi in linguaggio naturale secondo una logica di tipo lessicometrico propria della "Statistica Testuale", finalizzata all'analisi del contenuto di un testo. Tale approccio consente lo studio diretto di dati di natura linguistica mediante un'analisi automatica del testo. È particolarmente indicato per testi di ampie dimensioni, ossia di centinaia o migliaia di pagine. Taltac nato con l'idea non di sostituire il software già sviluppato per l'analisi di dati testuali sia in ambito statistico che linguistico, bensì di realizzare la più elevata integrazione possibile fra diversi livelli di analisi, sia lessicale che testuale. Convoglia in un unico ambiente misurazioni e risorse sia linguistiche, sia statistico-linguistiche ed è aperto all'utilizzo degli altri pacchetti già disponibili per la ricerca scientifica, in particolare quelli tipici dell'approccio lessicometrico (Spad, Lexico, Sphinx, Alceste, Hyperbase) e di quello lessico-grammaticale (Intex). Le diverse funzioni di Taltac costituiscono una preparazione indispensabile del testo per le analisi svolte in una logica di Text Analysis o di Text Mining.

⁷ <http://www.regione.emilia-romagna.it/terremoto/gli-atti-per-la-ricostruzione>

- Normalizzazione delle lessie, consistente nell'uniformare differenti forme grafiche abbreviate relative ad un unico significante, nel caso specifico in analisi ha riguardato principalmente le diverse abbreviature degli atti normativi citati nei testi;
- Trattamento Lessicale consistente: nell'individuazione delle unità di analisi lessicali (parole), intese sia come parole singole che come espressioni o locuzioni (multiword); nell'attribuzione delle meta-informazioni di tipo grammaticale e semantico alle “parole” costituenti il Vocabolario del corpus analizzato; nella selezione di dette unità su base quantitativa e/o probabilistica, in grado di estrarre le parole chiave, ovvero i termini significativi e peculiari;
- Definizione di un sistema ibrido lessico-testuale, per l'individuazione ed estrazione delle citazioni di atti normativi (strutture ad elementi variabili) presenti nei singoli documenti che formano il corpus;
- Classificazione Automatica, su base statistica (Cluster Analysis), consistente nel raggruppare i documenti secondo la loro massima similarità. In questo modo è possibile individuare a posteriori dei cluster (gruppi) di documenti fra loro omogenei in termini di tematizzazioni.
- Riclassificazione Fuzzy dei documenti, in grado di ottenere una categorizzazione plurima delle ordinanze in cui sono presenti più di una tematica. Si propone quindi il passaggio da un clustering non supervisionato ad un assegnazione multi-class supervisionata, attraverso la definizione di una misura del grado di appartenenza di ogni documento a ciascuna tipologia.

3. Corpus

Il Corpus sottoposto ad analisi è composto quindi dalle 349 Ordinanze emanate dal 08/06/2012 al 23/02/2015. L'ampiezza del linguaggio dopo il primo parsing del testo è di 31.969 forme grafiche originarie, ovvero le parole diverse o vocabolario del corpus (includendo anche le forme numeriche), mentre l'insieme dei dati testuali raccolti ammonta a 1.278.291 occorrenze di parole, equivalente a 2.550 pagine convenzionali di testo.

Ogni Ordinanza costituisce un'unità testuale del Corpus ed è stata suddivisa al suo interno in tre sezioni di testo: una prima parte “narrativa” di Premessa; una seconda parte di Disposizioni; una terza parte di Allegati.

Figura 2 - Consistenza in termini di parole ed occorrenze di ciascuna sezione di testo

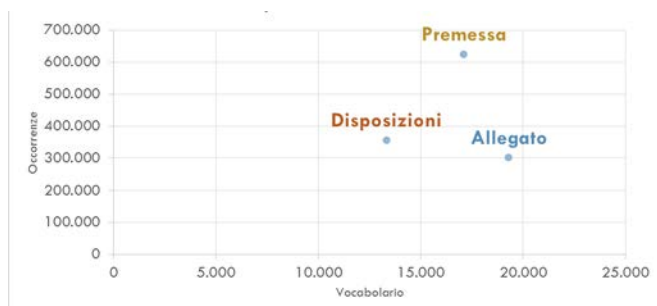


Figura 3 - Distribuzione trimestrale del numero di ordinanze emanate

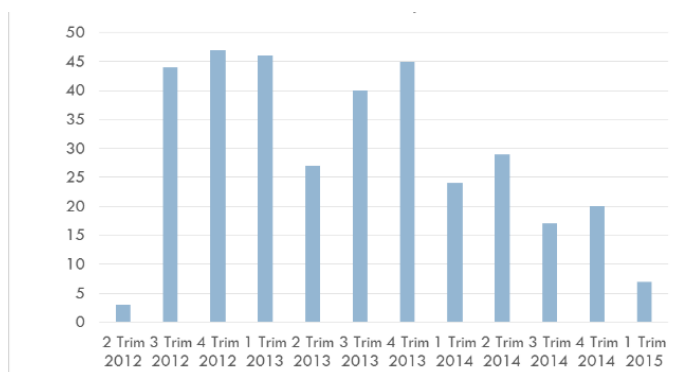
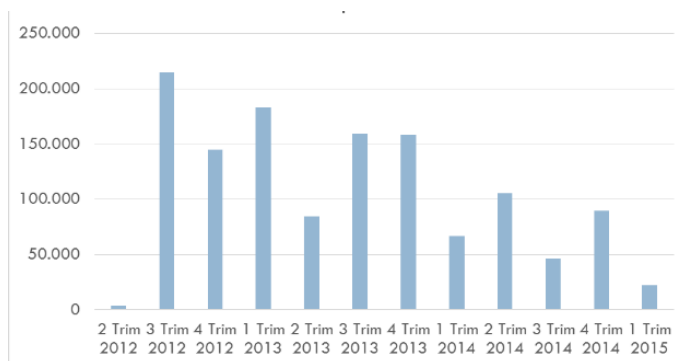


Figura 4 - distribuzione trimestrale delle occorrenze di parole (dimensioni dei testi) all'interno delle ordinanze emanate



4. Estrazione dell'informazione dalle ordinanze

4.1 Identificazione delle keyword

Generalmente nell'ambito dell'analisi automatica dei testi le forme di interesse sono rappresentate dalle parole chiave o Keyword, ovvero da termini che in un testo vengono selezionati secondo una funzione obiettivo. Il complesso di parole chiave di un documento costituisce un lessico particolarmente importante al fine dell'estrazione di informazione. In base agli indici utilizzati per estrarre le

keywords si possono ottenere di volta in volta differenti risultati. È possibile infatti estrarre da un testo differenti tipi di linguaggio: peculiare; specifico e rilevante.

Si dicono forme peculiari, per l'intero corpus, le unità che hanno un forte sovra/sotto uso rispetto ad un modello di riferimento. La peculiarità è misurata tramite lo scarto standardizzato della frequenza relativa dell'*i*-esima parola del Corpus rispetto alla frequenza normalizzata della stessa parola nel lessico (risorsa esogena) assunto come modello.

Per l'individuazione del linguaggio specifico si procede ad un calcolo di sovra-sotto utilizzo delle parole, ma a differenza del linguaggio peculiare non viene usata una risorsa esogena come modello di riferimento. La selezione delle parole specifiche viene effettuata rispetto alle caratteristiche endogene del corpus in analisi. Si misura quanto una parola risulti essere caratteristica o specifica di una parte del Corpus rispetto a tutte le altre parti. Dove per parti del Corpus si intendono le modalità di una variabile categoriale associata alla collezione dei testi. Ogni specificità positiva (sovra-utilizzo) di una parola equivale ad un uso superiore a quello "atteso" in una modalità, valutato rispetto al suo impiego medio nelle altre modalità che costituiscono la variabile categoriale. Per fare un esempio, è possibile individuare il linguaggio specifico degli uomini rispetto al linguaggio specifico delle donne, per un corpus in cui sia presente il *genere* come variabile categoriale.

Il linguaggio rilevante invece è calcolato a partire dall'indice TFIDF che permette di selezionare i termini maggiormente discriminanti. L'indice valorizza infatti le forme più frequenti, concentrate solo in pochi documenti.

Tabella 1 - Principali forme peculiari individuate

Linguaggio Peculiare			Verbi Peculiari		Verbi Sostantivati		
Forma grafica	Occorrenze totali	Scarto	Lemma	SCARTO	Forma grafica	Occorrenze	Lipe
sismici	2280	1005,081	ricomprendere	54,42942999	rendicontazione	369	294,0674
comma	4703	496,5274	provvedere	42,68231056	riparazione	1318	161,4169
recante	1265	437,5007	rideterminare	40,61265552	validazione	119	96,6489
modificazioni	2384	432,521	ravvisare	36,1195058	delocalizzazione	244	94,8002
sismico	1239	340,0461	avvalersi	34,24562443	urbanizzazione	385	92,8709
convertito	2824	336,9733	autorizzare	33,03039717	rimodulazione	227	74,8548
rendicontazione	369	294,0674	decorrere	28,0521069	ultimazione	163	67,08549
ordinanza	5044	285,3364	derogare	27,27561309	ricostruzione	1773	66,65686
commissariale	855	256,6566	effettuare	24,44978846	localizzazione	312	66,10748
Commissariale	693	207,9776	mitigare	22,24468663	prosecuzione	468	65,19921
sensi	2842	191,4409	emanare	21,88306527	realizzazione	1544	64,03517
allegato	1265	190,7202	individuare	21,8763639	liquefazione	135	60,80062
ordinanze	1134	162,7378	adibire	21,2898959	demolizione	400	57,6439
riparazione	1318	161,4169	allegare	19,62782154	attuazione	941	55,35149
urgenti	1871	154,2381	prorogare	19,46611235	erogazione	519	52,74117
interventi	5298	144,7885	riattivare	19,32843026	miglioramento	1061	49,98046
indifferibili	204	142,1019	disporre	18,64559285	esecuzione	1033	47,69091
agibilità	545	136,5553	rettificare	18,49807318	progettazione	534	46,45151
Modulari	264	132,9514	delocalizzare	17,94503349	espropriazione	106	38,95937

4.2 Identificazione delle Multiword

Attraverso l'operazione di tagging grammaticale delle forme del vocabolario è stato possibile estrarre delle multiword di tipo nominale utilizzando delle ER in

grado di ricercare le strutture sintattiche dei più comuni sintagmi nominali (Pavone, 2010). Si è proceduto pertanto ad individuare un elenco di sintagmi nominali dai quali sono stati selezionati 323 multiword di senso compiuto assunte come nuove entrate del Vocabolario e parte del lessico per le fasi successive dell'analisi.

Tabella 2 - Principali multiword riconosciute

Forma grafica	Occorrenze	Forma grafica	Occorrenze
eventi sismici	2084	parte integrante	436
decreto legge	2027	controllo preventivo	413
protezione civile	1145	enti attuatori	409
miglioramento sismico	775	opere pubbliche	389
Dipartimento della protezione civile	658	Bollettino Ufficiale Telematico	348
presidente del Consiglio dei ministri	511	rafforzamento locale	323
stato di emergenza	486	importo complessivo	323
beni culturali	456	attività di assistenza	309
Corte dei conti	450	Consiglio dei Ministri	302
ordinanza commissariale	449	delibera del Consiglio dei Ministri	298
interventi provvisionali	449	unità immobiliari	291
Agenzia regionale	448	opere di urbanizzazione	288
edifici scolastici	437	legge regionale	250

4.3 Il Cratere del sisma

Attraverso il riconoscimento dei Toponimi presenti all'interno del Corpus è stato possibile conteggiare le citazioni dei Comuni coinvolti all'interno delle Ordinanze, in modo da fornire l'intensità degli interventi nel Cratere del Sisma (Figura 1). È stato quindi possibile misurare anche le citazioni dei Comuni cronologicamente, sia su base annua che trimestrale⁸.

Figura 5 – Intensità delle citazioni dei comuni coinvolti nel cratere sismico all'interno delle ordinanze



⁸ Questa analisi viene ripresa nel lavoro di Piazza, Pagliacci e Russo (2015) sull'analisi cluster delle caratteristiche socio-economiche dei comuni dell'Emilia-Romagna.

5. Identificazione degli atti normativi citati nelle ordinanze

Attraverso il riconoscimento degli atti normativi citati all'interno delle Ordinanze si intende in un secondo momento costruire un'analisi di rete fra i diversi documenti. L'individuazione degli atti normativi nel Corpus avviene mediante la formalizzazione della loro struttura sintattica, nel caso specifico si tratta di una struttura ad elementi variabili, caratterizzata da elementi principali ed alcuni elementi opzionali.

Un atto normativo infatti è definito dalla sequenza di almeno due elementi fondamentali: la citazione dell'atto normativo generico (legge, decreto legge, ordinanza) e la data di emissione dell'atto. All'interno di questi due elementi è possibile trovare il numero dell'atto, oppure in alcuni casi il numero dell'atto può essere fuso con la data in un'unica forma grafica del tipo nn/anno (es. 74/2012).

Il processo di identificazione degli atti è caratterizzato da tre fasi: normalizzazione delle forme grafiche normative abbreviate; annotazione delle forme grafiche del vocabolario; esecuzione dell'espressione regolare

La normalizzazione delle forme grafiche abbreviate si rende necessaria soprattutto per l'eliminazione della punteggiatura che nel caso specifico non determina una separazione fra i caratteri ma al contrario indica una loro unione.

Tabella 3 - Esempio di normalizzazione della forma abbreviata di Decreto legislativo

Forma grafica	Occorrenze totali	Lunghezza	CAT	CAT_AC	CAT_SEM	Imprinting	Lemma
D.Lgs.	250	06	ERR				DLgs
D.Lgs	108	05	ERR				DLgs
d.lgs	41	05	ERR				DLgs
D.lgs	33	05	ERR				DLgs
D.lgs.	28	06	ERR				DLgs
d.lgs.	11	06	ERR				DLgs
d.lgsl.	8	07	ERR				DLgs
D.Lgs.vo.	3	09	ERR				DLgs
D.Lgs.vo	2	08	ERR				DLgs
D.Lgs.163/2006	2	14	ERR				DLgs 163/2006
D.Lgs.82/2005	2	13	ERR				DLgs 82/2005
d.Lgs	105		ERR				DLgs
d.Lgs.	106		ERR				DLgs
D.LGS.	106		ERR				DLgs

In un secondo momento si è definito un modello lessico-testuale per il riconoscimento nel corpus delle “date” citate nei testi nel formato classico /giorno/mese/anno/ e contestualmente si sono create delle apposite annotazioni (metainformazioni) alle forme grafiche relative agli atti normativi e le date. Nella tabella 4 sono riportati alcuni esempi di annotazioni all'interno del DB Vocabolario nel campo CATSEM.

Tabella 4 - Esempi di annotazioni nel db vocabolario

Forma grafica	Occorrenze totali	Lunghezza	CAT	CAT_AC	CAT_SEM
ordinanza	5.044	09	N	N	,AttoNorm,
legge	2.948	05	J	N+V	,AttoNorm,
DL	2.579	02			,AttoNorm,
Ordinanza	2.437	09	N	N	,AttoNorm,
decreto legge	2.197	13	N	N	,Altro,Polirematiche,AttoNorm,
74/2012	1.739	07			,nDATA,,DATA,
maggio	1.514	06	N	N	,Altro,Date abbreviate,mese,
L	1.252	01	N	N	,AttoNorm,
6 giugno 2012	1.217	13			,DATA,
Legge	993	05	J	N+V	,AttoNorm,
29 maggio 2012	979	14			,DATA,
122/2012	773	08			,nDATA,,DATA,
DLgs	746	04			,AttoNorm,
OCDPC	692	05			,AttoNorm,
decreto	635	07	J	N+V	,AttoNorm,
1 agosto 2012	599	13			,DATA,
Decreto	484	07	J	N+V	,AttoNorm,

A questo punto la ricerca degli specifici atti normativi avviene mediante la definizione di un'Espressione Regolare in cui si ricercano in sequenza all'interno del corpus le annotazioni effettuate alle forme del Vocabolario:

"CATSEM(AttoNorm) LAG6 CATSEM(DATA)"⁹

Tabella 5 – elenco delle principali Entità ricercate attraverso l'Espressione regolare

Etichette di riga	Somma
DL N 74/2012	1159
decreto legge 6 giugno 2012	601
legge 1 agosto 2012	444
LEGGE N 122/2012	434
DL 74/2012	409
DL 6 giugno 2012	253
OCDPC n 3/2012	251
legge 24 febbraio 1992	239
decreto legge 22 giugno 2012	201
L 24 novembre 2000	152
OCDPC n 1/2012	149

Le entità identificate tramite la ER sono riconosciute e lessicalizzate all'interno del DB Vocabolario. Vengono pertanto individuate 18.715 citazioni riferite ad un totale di 2.298 atti normativi diversi¹⁰.

Nella Tabella 6 vengono riportate le diverse tipologie di atti normativi citati mentre nella Tabella 7 sono elencate le definizioni dei principali atti normativi riconosciuti nel corpus.

⁹ Il comando LAG6 permette di catturare fino a 6 forme opzionali all'interno dei due elementi principali.

¹⁰ Nel caso delle citazioni di ordinanze si è verificato che l'espressione regolare abbia riconosciuto solamente il primo atto di un elenco, ad es. modifiche intervenute sulle ordinanze 29/2012, 51/2012 e 86/2012. Pertanto per il caso specifico delle citazioni di ordinanze per avere un conteggio corretto di tutte le ordinanze citate si è proceduto a riconoscere come tali anche le forme semplici del tipo nn/data non vincolate alla dicitura dell'atto normativo.

Tabella 6 - Tipologia di atti normativi citati nelle Ordinanze

Delibera dell'Assemblea Legislativa Regionale
 Decreto del Capo Dipartimento della Protezione Civile
 Decreto del Commissario Delegato
 Decreto del Presidente della Regione
 Delibera Assemblea Legislativa
 Delibera del Consiglio dei Ministri
 Delibera consiliare Castelfranco E.
 Delibera della Giunta Regionale
 Direttiva del Dipartimento della Protezione Civile
 Direttiva del Presidente del Consiglio dei Ministri
 Direttiva dell'Unione Europea
 Decreto-legge
 Decreto legislativo
 Decreto ministeriale
 Decreto del Presidente del Consiglio dei Ministri
 Decreto del Presidente della Repubblica
 Legge
 Legge regionale
 Ordinanza del Capo Dipartimento della Protezione Civile
 Ordinanza del Presidente del Consiglio dei Ministri
 Ordinanza del Commissario Delegato

Tabella 7 Lista dei titoli dei principali atti normativi citati

Atto normativo	N. citazioni
DECRETO-LEGGE 6 giugno 2012, n. 74. Interventi urgenti in favore delle popolazioni colpite dagli eventi sismici che hanno interessato il territorio delle province di Bologna, Modena, Ferrara, Mantova, Reggio Emilia e Rovigo, il 20 e il 29 maggio 2012.	2.878
LEGGE 1 agosto 2012, n. 122. Conversione in legge, con modificazioni, del decreto-legge 6 giugno 2012, n. 74, recante interventi urgenti in favore delle popolazioni colpite dagli eventi sismici che hanno interessato il territorio delle province di Bologna, Modena, Ferrara, Mantova, Reggio Emilia e Rovigo, il 20 e il 29 maggio 2012.	1.436
DECRETO-LEGGE 22 giugno 2012, n. 83. Misure urgenti per la crescita del Paese.	410
LEGGE 7 agosto 2012, n. 134. Conversione in legge, con modificazioni, del decreto-legge 22 giugno 2012, n. 83, recante misure urgenti per la crescita del Paese.	395
LEGGE 24 novembre 2000, n. 340 Disposizioni per la delegificazione di norme e per la semplificazione di procedimenti amministrativi - Legge di semplificazione 1999.	389
DECRETO LEGISLATIVO 12 aprile 2006, n. 163 Codice dei contratti pubblici relativi a lavori, servizi e forniture in attuazione delle direttive 2004/17/CE e 2004/18/CE.	318
Ordinanza n. 57 del 12 ottobre 2012 - Criteri e modalità per il riconoscimento dei danni e la concessione dei contributi per la riparazione, il ripristino, la ricostruzione di immobili ad uso produttivo, per la riparazione e il riacquisto di beni mobili strumentali all'attività, per la ricostituzione delle scorte e dei prodotti e per la delocalizzazione, in relazione agli eventi sismici del 20 e 29 maggio 2012	291
LEGGE 24 febbraio 1992, n. 225. Istituzione del Servizio nazionale della protezione civile.	276
DELIBERAZIONE DEL CONSIGLIO DEI MINISTRI 22 maggio 2012 Dichiarazione dello stato di emergenza in conseguenza degli eventi sismici che hanno colpito il territorio delle province di Bologna, Modena, Ferrara e Mantova il giorno 20 maggio 2012.	270
Ordinanze del Capo Dipartimento della Protezione Civile n. 3 del 2 giugno 2012: interventi urgenti di protezione civile conseguenti agli eventi sismici che hanno colpito il territorio delle province di Bologna, Modena, Ferrara, Reggio Emilia, Mantova e Rovigo nel mese di maggio 2012 Pubblicata nella Gazzetta	258

Ufficiale n. 130 del 6 giugno 2012	
Ordinanza n. 86 del 6 dicembre 2012 - Criteri e modalità di assegnazione di contributi per la riparazione, il ripristino con miglioramento sismico o la demolizione e ricostruzione di edifici e unità immobiliari ad uso abitativo che hanno subito danni gravi a seguito degli eventi sismici del 20 e 29 maggio 2012 e che sono stati dichiarati inagibili (ESITO E1, E2 o E3)	240
Ordinanza del Capo Dipartimento della Protezione Civile n. 1 del 22 maggio 2012: primi interventi urgenti per gli eventi sismici che hanno colpito le province di Bologna, Modena, Ferrara e Mantova il 20 maggio 2012	238
Ordinanza n. 29 del 28 agosto 2012 - Criteri e modalità di assegnazione di contributi per la riparazione e il ripristino immediato di edifici e unità immobiliari ad uso abitativo danneggiati dagli eventi sismici del 20 e 29 maggio 2012 e temporaneamente o parzialmente inagibili.	208
Ordinanza n. 51 del 5 ottobre 2012 - Ordinanza n. 51/2012 coordinata con le modifiche apportate a seguito della firma del Protocollo d'intesa firmato dal Ministro dell'Economia e delle Finanze e dai Presidenti delle Regioni Emilia-Romagna, Lombardia e Veneto in qualità di Commissari delegati in data 4 ottobre 2012 pubblicato sulla GU del 24 ottobre 2012 e dell'approvazione del D.L. 174 del 10 ottobre 2012	187
DECRETO-LEGGE 26 aprile 2013, n. 43 Disposizioni urgenti per il rilancio dell'area industriale di Piombino, di contrasto ad emergenze ambientali, in favore delle zone terremotate del maggio 2012 e per accelerare la ricostruzione in Abruzzo e la realizzazione degli interventi per Expo 2015.	177

Fra gli atti normativi più citati risaltano le ordinanze riferite ai *Criteri e modalità di assegnazione dei contributi*, fra queste la più citata risulta essere l'ordinanza commissariale 57/2012¹¹ con 291 occorrenze. L'elenco completo dei 2.298 atti normativi riconosciuti sarà utilizzato in una seconda fase dell'analisi per la creazione di una matrice *Ordinanze × Atti normativi* sulla quale generare un'analisi di rete che metta in luce quali ambiti normativi sono associati alle ordinanze emanate per far fronte all'emergenza e alla ricostruzione e anche in questo modo contribuire alla definizione delle filiere normative rilevanti in fase di stesura della nuova legge in discussione al Senato.

6. Individuazione delle tematiche delle ordinanze

Come accennato nel paragrafo introduttivo uno degli aspetti della ricerca in ambito testuale mira a risolvere i problemi relativi alla classificazione e categorizzazione automatica dei dati. Generalmente in assenza di categorie predefinite dal ricercatore, si procede ad una clusterizzazione dei documenti in base alla loro massima similarità, permettendo in questo modo di individuare a posteriori le tematizzazioni prevalenti. Nella maggior parte delle procedure di analisi dei dati testuali, una classificazione di questo tipo porta a classi disgiunte di documenti, mentre i documenti appartenenti a diversi clusters spesso hanno elementi o tratti semantici in comune. La prima circostanza porta a classificazioni univoche, la seconda a classificazioni di tipo fuzzy (Zadeh, 1977). Nel paragrafo 6.1 vengono

¹¹ Ordinanza 57/2012 - Criteri e modalità per il riconoscimento dei danni e la concessione dei contributi per la riparazione, il ripristino, la ricostruzione di immobili ad uso produttivo, per la riparazione e il riacquisto di beni mobili strumentali all'attività, per la ricostituzione delle scorte e dei prodotti e per la delocalizzazione, in relazione agli eventi sismici del 20 e 29 maggio 2012

esposti i risultati di una classificazione disgiunta non supervisionata delle ordinanze mentre nel paragrafo 6.2 si illustrano i risultati di una assegnazione multi-classe supervisionata.

6.1 Clustering non supervisionato

Questa fase dell'elaborazione consiste nel raggruppare i documenti in gruppi o classi (clusters) sulla base della similarità della distribuzione dei termini selezionati nelle precedenti fasi dell'analisi. Questa fase di clustering sulla sola base di keywords costituisce una classificazione non supervisionata ed univoca dei documenti, che riflette la similarità a livello semantico esistente tra i documenti così profilati. Tale omogeneità concettuale espliciterà il tema o tratto semantico prevalente in ciascun gruppo di documenti individuato, riassumibile in una categoria non definita a priori ma ottenuta mediante l'analisi.

A tale scopo si considera una matrice del tipo Documenti x Keywords ($n \times p$):

	Keyword ₁	...	Keyword _j	...	Keyword _p
Doc ₁	x ₁₁	...	x _{1j}	...	x _{1K}
...
Doc _i	x _{il}	...	x _{ij}	...	x _{iK}
...
Doc _n	x _{nl}	...	x _{nj}	...	x _{nK}

La matrice profila i diversi documenti da classificare con le occorrenze dei termini rilevanti in essi contenute. Sottoponendo tale matrice alla classica filiera di analisi statistica multidimensionale, “analisi fattoriale + cluster analysis”, si otterrà una partizione in K gruppi di documenti, omogenei al loro interno ed eterogenei fra loro.

I K gruppi di documenti ottenuti rappresentano classi disgiunte e i lessici associati a ciascun gruppo definiscono il campo semantico o “tema” della classe. Il campo semantico, espressione dell'omogeneità dei gruppi, è visualizzabile sul piano fattoriale attraverso la vicinanza delle forme che lo costituiscono.

Attraverso l'Analisi Fattoriale delle Corrispondenze (AC) il complesso spazio multidimensionale delle variabili viene ridotto a pochi fattori principali in grado rappresentare, di volta in volta su dimensioni chiamate assi fattoriali, le relazioni fra gli elementi della matrice dei dati¹². L'AC produce infatti la miglior rappresentazione simultanea dei punti profili riga (vs profili colonna) in ciascun piano fattoriale, e prima ancora su ciascuno dei suoi assi (Bolasco, 1999).

L'interpretazione degli assi fattoriali può avvenire unicamente visualizzando sui piani, oppure osservando i tabulati, i campi semantici che contribuiscono alla variabilità di ciascun fattore. Infatti, dal posizionamento delle parole nei sottospazi fattoriali è possibile dedurre le relazioni significative fra le unità lessicali: la lo-

¹² Cfr. (Benzecri, 1973) (Benzecri, 1981) (Lebart & Salem, 1988)

ro vicinanza sul piano è funzione dell'associazione delle loro occorrenze nei documenti. Ogni rappresentazione dell'insieme dei punti delle righe (documenti) o delle colonne (unità lessicali) della matrice rappresenta una mappa delle relazioni fra di essi¹³.

Sulla base delle correlazioni fra documenti e parole spiegate dai diversi assi sui piani fattoriali, è possibile effettuare una Cluster Analysis in modo da definire dei gruppi di documenti. La procedura di Cluster Analysis in sequenza all'analisi delle corrispondenze, utilizza il metodo di classificazione gerarchica che segue il criterio di aggregazione di Ward.

Si genera pertanto una matrice *Ordinanze x Keywords*¹⁴ (349 x 4.273) da sottoporre ad Analisi delle Corrispondenze + Cluster Analysis¹⁵. Nello specifico si è optato per considerare per ciascun ordinanza esclusivamente il testo delle Premesse e delle Disposizioni delle, escludendo la sezione Allegati. La scelta è motivata dal fatto di voler individuare le tematiche descrittive delle azioni intraprese dai Commissari Delegati escludendo i possibili elementi di *rumore*¹⁶.

Nelle figure 5 e 6 sono rappresentate le nuvole dei punti rispettivamente della distribuzione dei documenti e delle parole sui piani fattoriali f1-f2. La sovrapposizione grafica dei due piani fattoriali rappresenta la correlazione fra i docu-

¹³ Ogni dimensione fattoriale rappresenta un continuum lungo il quale si dispongono gli elementi oggetto di studio. Le coordinate dei punti rappresentano delle misure di scarto dal baricentro e quindi l'allontanamento da esso intende una diversa e opposta polarizzazione del punto di vista che tale asse fattoriale rappresenta del fenomeno. Queste caratteristiche permettono di interpretare gli assi fattoriali in base ai punti che rappresentano. Più i punti sono distanti dall'origine più sono significativi. Inoltre in base al posizionamento delle parole nei sottospazi fattoriali dell'AC è possibile dedurre le relazioni significative fra le unità lessicali. La loro vicinanza su un asse e ancor più su un piano, è funzione dell'associazione delle loro occorrenze nei documenti. Le parole vicine su un piano fattoriale indicheranno una loro co-occorrenza all'interno di determinati documenti. Allo stesso modo documenti vicini sul piano fattoriale indicheranno una similitudine in termini del loro lessico.

¹⁴ Le 4.273 keywords, della matrice sottoposta ad analisi fattoriale, sono state selezionate nelle precedenti fasi di analisi lessicale del Vocabolario in base ai valori del linguaggio peculiare e in base alle caratteristiche grammaticali, selezionando unicamente le forme annotate come sostantivi, aggettivi e verbi.

¹⁵ L'analisi delle corrispondenze e la cluster analysis sono state eseguite con l'utilizzo del programma SPAD (Système portable pour l'analyse des données) (Lebart & Morineau, 1984), distribuito dal Cisia di Parigi - Centre de statistique et d'informatique appliquées - è stato sviluppato da un gruppo di autori esponenti della scuola francese dell'Analyse des données.

Si tratta di un software statistico modulare in grado di importare matrici di dati sotto forma di Base Spad alla quale è poi possibile applicare un insieme di metodi statistici come: analisi fattoriali, classificazioni, archiviazioni, segmentazioni ed analisi testuale. Nell'approccio di Spad le tecniche di classificazione automatica e quelle fattoriali sono complementari. In primo luogo si effettua un'analisi delle corrispondenze in grado di ridurre il numero delle variabili condensandole in pochi fattori sintetici. Successivamente sulla base dei fattori così ottenuti si effettua una classificazione automatica dei casi riducendoli a una tipologia dove ciascun tipo è costituito da un insieme di individui molto simili fra di loro rispetto all'insieme delle caratteristiche che formano la tipologia.

¹⁶ Non tutte le ordinanze infatti sono dotate di allegati e il più delle volte gli allegati sono costituiti da tabelle oppure da documenti integrali di Gazzette Ufficiali o Testi Coordinati.

menti e il lessico analizzato. In un'ottica di analisi testuale infatti l'analisi fattoriale permette di studiare le relazioni tra le keywords nello spazio definito dai documenti (o le relazioni tra documenti nello spazio definito dalle keywords).

Osservando la distribuzione documenti e delle partizioni sul piano fattoriale nella Figura 5 è possibile fare delle prime considerazioni sul risultato.

Si contraddistingue un gruppo centrale *baricentrico* (Cluster 2) da cui si separano 3 code di documenti che vanno a generare in maniera quasi naturale 3 diversi raggruppamenti di ordinanze molto polarizzati fra di loro.

Figura 6 - Piano fattoriale f1f2 – Distribuzione dei: documenti, partizioni, involuipi convessi

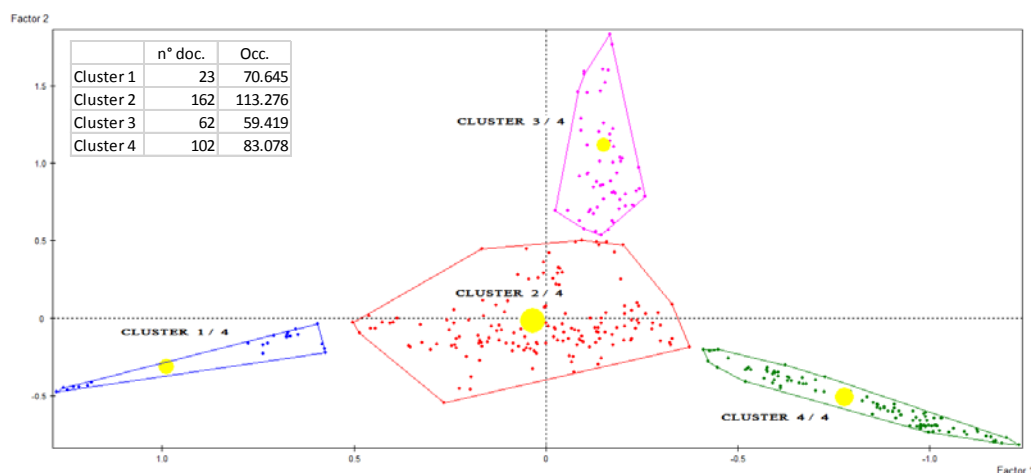
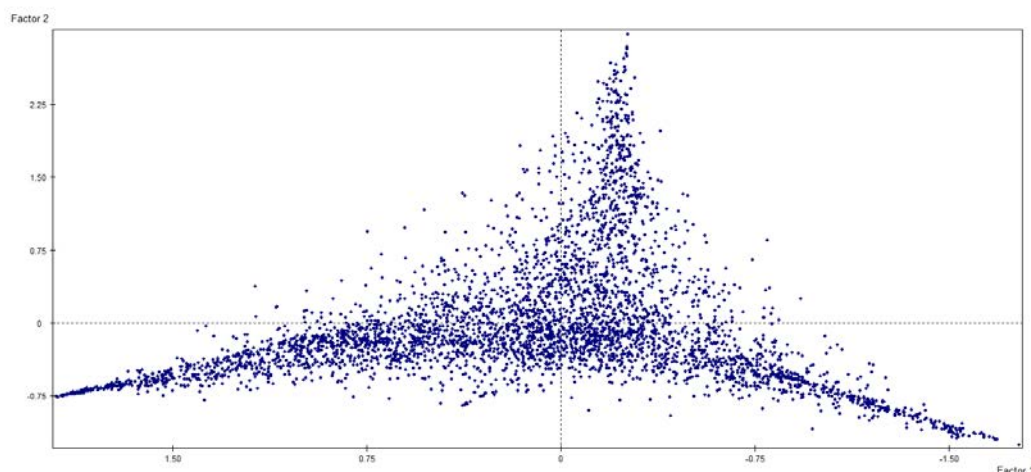


Figura 7 – Piano fattoriale f1f2 - Distribuzione delle keywords



L'asse fattoriale 1 spiega la contrapposizione tra il Cluster1 rispetto al Cluster4, mentre il fattore 2 evidenzia maggiormente la polarizzazione fra il Cluster3 rispetto al Cluster4.

Il Cluster1, risulta essere formato da pochi documenti (23) caratterizzati da un gran numero di occorrenze (70.645), inoltre all'interno dello stesso gruppo si

evidenzia la formazione di due sotto-gruppi di documenti. In particolare si distaccano fra loro 7 Ordinanze, sul lato sinistro del gruppo. Questo piccolo sottogruppo di documenti è molto concentrato, i punti sono molto vicini fra di loro indicando una fortissima similarità dei contenuti.

Andando a fare una verifica mirata di tali documenti risulta che pur non trattandosi di Testi Coordinati, nelle suddette ordinanze vengono ripetuti interi paragrafi dell'Ordinanza 57/2012 con le modifiche di volta in volta apportate. Il dettaglio dei titoli delle 7 ordinanze concentrate nel cluster 1 è riportato nella Tabella 8.

Tabella 8 - Dettaglio delle 7 ordinanze del Cluster 1/4

Ordinanza_42_2013.03.29	Modifiche all'Ordinanza n. 57 del 12 ottobre 2012 come modificata dall'Ordinanza n. 64 del 29 ottobre 2012, dall'Ordinanza n. 74 del 15 novembre 2012 e dall'Ordinanza n. 15 del 15 febbraio 2013. Criteri e modalità per il riconoscimento dei danni e la concessione dei contributi per la riparazione, il ripristino, la ricostruzione di immobili ad uso produttivo, per la riparazione e il riacquisto di beni mobili strumentali all'attività, per la ricostituzione delle scorte e dei prodotti e per la delocalizzazione, in relazione agli eventi sismici del 20 e 29 maggio 2012. Approvazione delle Linee Guida per la presentazione delle domande e le richieste di erogazione dei contributi
Ordinanza_113_2013.09.30	Modifiche all'Ordinanza n. 57 del 12 ottobre 2012 come modificata dall'Ordinanza n. 64 del 29 ottobre 2012, dall'Ordinanza n. 74 del 15 novembre 2012, dall'Ordinanza n. 15 del 15 febbraio 2013 e dall'Ordinanza n. 42 del 29 marzo 2013 Criteri e modalità per il riconoscimento dei danni e la concessione dei contributi per la riparazione, il ripristino, la ricostruzione di immobili ad uso produttivo, per la riparazione e il riacquisto di beni mobili strumentali all'attività, per la ricostituzione delle scorte e dei prodotti e per la delocalizzazione, in relazione agli eventi sismici del 20 e 29 maggio 2012. Approvazione delle Linee Guida per la presentazione delle domande e le richieste di erogazione dei contributi (registrata alla Corte dei Conti Sezione Regionale di Controllo per l'Emilia-Romagna, in data 02 ottobre 2013 registro n. 1 foglio n. 367).
Ordinanza_70_2014.10.16	Modifiche all'Ordinanza n. 57 del 12 ottobre 2012 come modificata dall'Ordinanza n. 64 del 29 ottobre 2012, dall'Ordinanza n. 74 del 15 novembre 2012, dall'Ordinanza n. 15 del 15 febbraio 2013, dall'Ordinanza n. 42 del 29 marzo 2013, dall'Ordinanza n. 113 del 30 settembre 2013, dall'Ordinanza n. 3 del 27 gennaio e dall'Ordinanza n. 28 del 17 aprile 2014 Criteri e modalità per il riconoscimento dei danni e la concessione dei contributi per la riparazione, il ripristino, la ricostruzione di immobili ad uso produttivo, per la riparazione e il riacquisto di beni mobili strumentali all'attività, per la ricostituzione delle scorte e dei prodotti e per la delocalizzazione, in relazione agli eventi sismici del 20 e 29 maggio 2012. Approvazione delle Linee Guida per la presentazione delle domande e le richieste di erogazione dei contributi.
Ordinanza_74_2012.11.14	Modifiche all'ordinanza commissariale n. 57 del 12 ottobre 2012 "Criteri e modalità per il riconoscimento dei danni e la concessione dei contributi per la riparazione, il ripristino, la ricostruzione di immobili ad uso produttivo, per la riparazione e il riacquisto di beni mobili strumentali all'attività, per la ricostituzione delle scorte e dei prodotti e per la delocalizzazione, in relazione agli eventi sismici del 20 e 29 maggio 2012. Approvazione delle linee guida per la presentazione delle domande e le richieste di erogazione dei contributi previsti dall'Ordinanza n.57/2012
Ordinanza_28_2014.04.17	Modifiche all'Ordinanza n. 57 del 12 ottobre 2012 come modificata dall'Ordinanza n. 64 del 29 ottobre 2012, dall'Ordinanza n. 74 del 15 novembre 2012, dall'Ordinanza n. 15 del 15 febbraio 2013, dall'Ordinanza n. 42 del 29 marzo 2013 e dall'Ordinanza n. 113 del 30 settembre 2013.
Ordinanza_3_2014.01.27	Modifiche all'Ordinanza n. 57 del 12 ottobre 2012 come modificata dall'Ordinanza n. 64 del 29 ottobre 2012, dall'Ordinanza n. 74 del 15 novembre 2012, dall'Ordinanza n. 15 del 15 febbraio 2013, dall'Ordinanza n. 42 del 29 marzo 2013 e dall'Ordinanza n. 113 del 30 settembre 2013.
Ordinanza_15_2013.02.15	Modifiche all'Ordinanza n. 57 del 12 ottobre 2012 come modificata dall'Ordinanza n. 64 del 29 ottobre 2012 e dall'Ordinanza n. 74 del 15 novembre 2012 Criteri e modalità per il riconoscimento dei danni e la concessione dei contributi per la riparazione, il ripristino, la ricostruzione di immobili ad uso produttivo, per la riparazione e il riacquisto di beni mobili strumentali all'attività, per la ricostituzione delle scorte e dei prodotti e per la delocalizzazione, in relazione agli eventi sismici del 20 e 29 maggio 2012. Approvazione delle Linee Guida per la presentazione delle domande e le richieste di erogazione dei contributi

Dall'osservazione dei lessici associati a ciascun gruppo di documenti è possibile invece individuare il campo semantico rappresentativo di ciascun gruppo di ordinanze e stabilire a posteriori la tematica che definisce ciascuna classe. In questo modo è possibile inoltre interpretare le differenti polarizzazioni rappresentate sugli assi fattoriali.

Di seguito sono elencati i dettagli dei 4 clusters con le relative tematiche e alcune delle unità lessicali che caratterizzano ciascun gruppo:

Cluster 1 (23 documenti – 70.645 occ.) – **Criteri Concessione Contributi** – contributo, domanda, edificio, testo, progetto, superficie, edifici, domanda di contributo, miglioramento sismico, rafforzamento locale, costo dell'intervento, costo convenzionale, costo ammissibile, istituto di credito, unità immobiliare, relazione geologica, risarcimento, demolizione, immobile, richiedente, impresa affidataria, proprietari;

Cluster 2 (162 documenti – 113.276 occ.) – **Gestione Assegnazione Risorse** – convenzione, assegnazione delle risorse, gestione, fondi, risorse, conto capitale, somministrazione, stanziamento, affidamento, monitoraggio, documentazione antimafia, leggi antimafia, erogazione, graduatorie, risorse assegnate, svolgimento, accertamento, agevolazioni, MUDE, procedure informatiche, espletamento, supporto, concessione, opere pubbliche, lavori pubblici, beni culturali, Sistemazione, riparto, Manutenzione, smaltimento, asportazione, bonifica, condizioni di sicurezza, assunzione di personale, territorio, destinatario, assegnatario, nuclei familiari, Nucleo di Valutazione, Amministrazioni, Comuni, Chiese, edilizia scolastica, alloggi, edifici di culto;

Cluster 3 (62 documenti – 59.419 occ.) – **Opere Urgenti Scuole Municipi Chiese** – Edifici scolastici, anno scolastico, Scuole, aree, opere di urbanizzazione, riparazione immediata, prefabbricati modulari scolastici, riapertura delle scuole, Edifici Pubblici Temporanei, realizzazione, localizzazione, Programma Operativo Municipi, Lavori, temporanee, celerità, urgenza, Prefabbricati Modulari, contributo integrativo, riparazione immediata, Costruzione, appalto, municipi, straordinaria, Contratti, manutenzione ordinaria, integrativo, esito di agibilità, riparate, strutture scolastiche, locazione, scuole paritarie, edifici municipali, prefabbricati modulari municipali, edifici religiosi, chiese temporanee, esercizio del culto, espropriazione, collaudo tecnico;

Cluster 4 (102 documenti – 83.078 occ.) – **Interventi Assistenza Popolazione** – interventi provvisori, urgente, oneri finanziari, protezione civile, vigili del fuoco, stato di emergenza, amministrazioni locali, oneri di natura sociale, interventi di soccorso, assistenza, popolazione, incolumità, anziani, provvisoria, disabili NO autosufficienti, accoglienza, contabilità speciale, sanitarie, coordinamento operativo, strutture ricettive, assistenza specialistica, volontariato, salvaguardia, strutture sociosanitarie, attività di assistenza, qualità, campi.

Da un punto di vista semantico è ora possibile dare un'interpretazione degli assi fattoriali grazie alla caratterizzazione tematica dei gruppi di documenti. L'asse fattoriale 1 contrappone infatti in maniera netta i **Criteri Concessione Contributi** rispetto agli **Interventi di Assistenza alla popolazione**, in questo modo il fattore 1 spiega il passaggio da un aspetto teorico di definizione dei criteri ad verso un aspetto pratico degli interventi urgenti da intraprendere. Allo stesso modo la contrapposizione sul fattore 2 delle **Opere Urgenti per Scuole, Municipi e Chiese** rispetto agli **Interventi di Assistenza alla popolazione** sta ad indicare, all'interno dell'ambito dell'urgenza degli interventi, il passaggio fra il ripristino dei luoghi

fondamentali della vita pubblica rispetto agli interventi di aiuto diretto alla popolazione.

Sempre dall'osservazione della Figura 6 risulta evidente la vicinanza di documenti lungo le linee di confine definite dagli involuipi convessi. Ciò sta ad indicare che ordinanze classificate in maniera disgiunta con la Cluster Analysis hanno tratti semantici comuni. Per ottenere una più accurata categorizzazione dei documenti analizzati, nel paragrafo successivo viene illustrata una procedura di riclassificazione di tipo fuzzy¹⁷ delle ordinanze rispetto alle tematiche, in modo da ottenere una loro categorizzazione plurima.

6.2 Categorizzazione plurima supervisionata

Il criterio utilizzato per ottenere la classificazione multipla è basato sul calcolo dell'indice TFIDF¹⁸ (nel seguito TI) che attribuisce un peso alle parole considerando non solo l'importanza relativa di una forma in un documento, ma anche il suo potere di discriminazione rispetto all'intero corpus (Balbi & Misuraca, 2005).

Per il nostro scopo, si considerano i quattro dizionari tematici derivanti dalla cluster analysis, come altrettante queries testuali rispetto alle quali calcolare il TI per ciascun documento¹⁹. Tali TI sono gli elementi interni di una matrice Documenti x Queries (349x4). In questo modo i valori dell'indice TI, calcolati per ciascun lessico tematico rispetto a ciascuna ordinanza, ci consentono di misurare la pertinenza di un documento rispetto alle quattro tematiche. Ad esempio ordinando le ordinanze per valore decrescente di TI rispetto alla tematica **Criteri Concessione Contributi**, avremo un ordinamento di pertinenza dei documenti rispetto a tale query (Bolasco & Pavone, 2008). Per ottenere il valore *fuzzy* di appartenenza di un documento rispetto ad una *query* (tematica) si procede a relativizzare i valori di TI rispetto al totale dei 4 *scores* per ciascun documento. In questo modo otteniamo il grado di appartenenza del documento rispetto a ciascuna *query*. La nuova matri-

¹⁷ A differenza della logica disgiuntiva in cui un elemento x appartiene ($I(A,x)=1$) o non appartiene ($I(A,x)=0$) ad un insieme A , nella logica fuzzy un elemento x appartiene a un insieme A con un grado di "verità" compreso tra 0 e 1. Tale grado di appartenenza è denominato "membership" $m(A,x)$ e sostituisce l'indicatore $I(A,x)$: $0 \leq m(A,x) \leq 1$

Con grado di appartenenza si intende quanto è vera una proprietà: questa può essere oltre che vera (= al valore 1) o falsa (= al valore 0), anche sfumata, ovvero pari a un valore intermedio (ZADEH, 1977)

¹⁸ Il TI è un indice tipico dell'information retrieval e comunemente si usa per ordinare secondo una misura di rilevanza i documenti recuperati sulla base di una query, come generalmente accade nei motori di ricerca sul web. Il TI attribuisce infatti un peso ad ogni forma del vocabolario sulla base della loro frequenza nei singoli documenti attribuendogli una maggior importanza se tale frequenza è concentrata solo in pochi documenti. L'indice si articola in due componenti:

TF = Term Frequency; IDF = Inverse Document Frequency

¹⁹ In particolare il software Taltac restituisce un valore di TI normalizzato in base al frammento e standardizzato rispetto al valore massimo di TI di ciascuna query.

ce $D \times Q$ conterrà quindi i valori di appartenenza fuzzy dei documenti nelle le tematiche rappresentate dai lessici definiti con le query testuali. Una volta individuati i gradi di appartenenza di ciascun Ordinanza rispetto a ciascuna tematica si procede alla categorizzazione plurima dei documenti. Presupponendo quindi un'equidistribuzione delle tematiche all'interno dei documenti, convenzionalmente viene assegnata la categoria di appartenenza ad un tema ogniqualvolta il grado di appartenenza supera il livello di equidistribuzione, che nel caso in analisi di quattro categorie equivale a 0,25. Nella Tabella 9 viene riportato un riepilogo delle categorizzazioni disgiunte e plurime.

Tabella 9 - Classificazione delle 349 ordinanze in categorie univoche e plurime (fuzzy)

categorie disgiunte		categorie plurime (fuzzy)				
		CAT-1	CAT-2	CAT-3	CAT-4	
Criteria concessione contributi	23	CAT-1	20	9	0	0
Gestione Assegnazione Risorse	162	CAT-2	9	137	14	8
Opere Urgenti Scuole Municipi Chiese	62	CAT-3	0	14	60	0
Interventi Assistenza Popolazione	102	CAT-4	0	8	0	100
totale	349					

Nel complesso, per 33 su 349 ordinanze vi sono sovrapposizioni tra le tematiche individuate nell'analisi cluster. Si tratta di sovrapposizione che insistono sul Cluster 2-Gestione Assegnazione Risorse, che nel piano fattoriale risulta quello centrale, e quindi mediamente rappresentativo di tematiche presenti nel complesso del corpus di ordinanze²⁰. Le sovrapposizioni incidono maggiormente sul Cluster 1- Criteri Concessione Contributi, il più piccolo cluster individuato, che ha quai un terzo delle ordinanze in comune con il cluster 2, minore è l'importanza quantitativa per gli altri cluster.

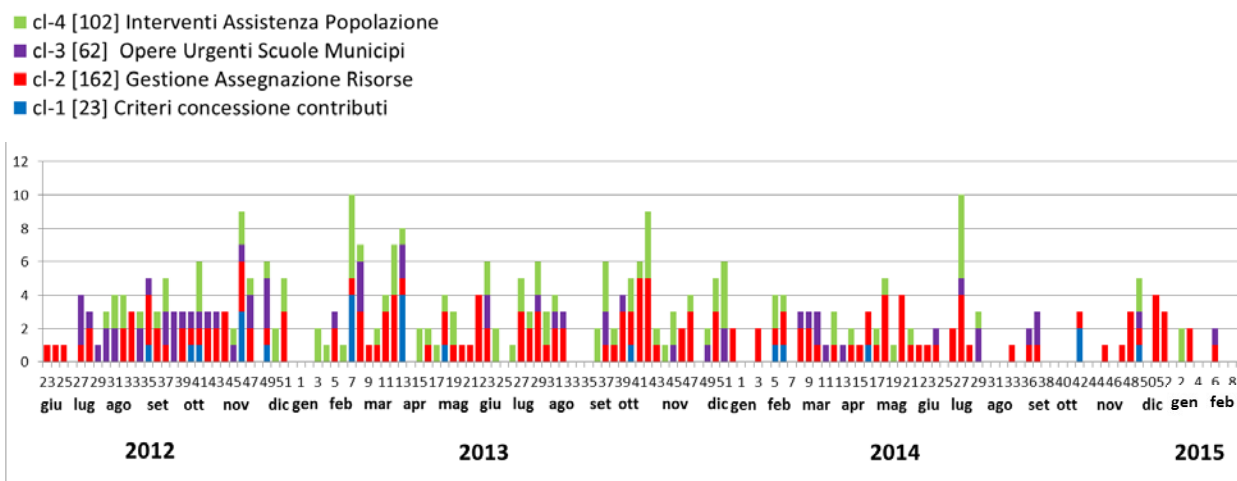
7. Analisi temporale dei temi delle ordinanze

La Figura 8 ci consente di leggere, nella scansione settimanale della data in cui sono state emanate, le ordinanze che avevamo rappresentato nella Figura 1 su

²⁰ Solamente in un caso la categorizzazione plurima ha classificato un'ordinanza in tre tematiche. Si tratta dell'ordinanza n. 2 del 16/06/2012, a cui è stata attribuita l'appartenenza alle tematiche 2, 3 e 4. Leggendo il testo di tale ordinanza, si è riscontrata la presenza di forme grafiche, come *capo dipartimento della protezione civile, urgente, lesioni, opere provvisionali*, che caratterizzano il lessico della tematica 4. Mediamente l'80% delle occorrenze di tali parole ricorre infatti all'interno della tematica *assistenza alla popolazione*, ma nel caso specifico dell'ordinanza n.2/2012, il richiamo alla protezione civile, ad esempio, fa riferimento alle prime ordinanze emanate da tale organo per gestire gli interventi post-sisma. Pertanto la presenza di tali forme grafiche all'interno dell'ordinanza genera un *falso positivo* che si è provveduto a correggere manualmente eliminando per il documento l'appartenenza alla tematica 4.

base mensile. La classificazione delle ordinanze per tematica ci aiuta ora a delineare la dinamica temporale delle aree di intervento nel periodo in esame.

Figura 8 - Ordinanze per cluster di assegnazione e mese in cui sono state emanate, 8.06.2012-20.02.2015



Le ordinanze che riguardano la gestione e assegnazione delle risorse [cl-2], il gruppo più grande costituito da 162 ordinanze, sono le prime ad essere emanate e continuano ad essere emanate lungo tutto il periodo considerato (a meno di poche interruzioni). Dopo il termine della gestione direttamente a capo della Protezione Civile, dopo qualche settimana prende avvio la sequenza di ordinanze che riguardano interventi di assistenza alla popolazione [cl-4]; seguono in modo regolare per i primi mesi, ma continuano nel corso dei tre anni, ad essere emanate ordinanze su opere urgenti per scuole, municipi e chiese [cl-3]: sono ordinanze che strutturano il passaggio dall'emergenza alla ricostruzione e che delineano le condizioni per rendere possibile la ripresa dell'anno scolastico (con il piano di ricostruzione delle scuole) e il ripristino dei luoghi di governo del territorio (i municipi) e di culto (le chiese). Infine, a punteggiare per l'intero periodo, ci sono le ordinanze sui criteri di concessione dei contributi [cl-1]: sono il gruppo più piccolo del nostro campo di osservazione (29 ordinanze).

Conclusioni

Quello che abbiamo tracciato è un quadro preliminare di analisi del corpus di testi da cui si dipaneranno tre sviluppi.

Il primo riguarda la rete delle citazioni degli atti normativi elaborata attraverso lo specifico modello lessico testuale descritto nella sezione 5. Una volta estratte dal corpus di testi in esame le citazioni di normative nazionali e di altre ordinanze, i gruppi di citazioni vengono elaborati con un algoritmo di analisi delle reti (Infomap) sulla base di teoria dell'informazione (Rosvall & Bergstrom, 2008;

De Domenico et al, 2015). Questa analisi (Pavone, Righi e Russo 2015) mette in evidenza quali sono i sottoinsiemi di nodi più intensamente connessi e consente di delineare quali sono e come si intrecciano tra di loro le questioni più rilevanti coinvolti nelle aree di intervento dopo un disastro naturale.

Il secondo filone su cui si sta sviluppando la ricerca riguarda l'analisi temprale di modifiche e integrazioni delle ordinanze, con riferimento ai diversi temi e alle specifiche tipologie di modifiche che sono state necessarie (Palmirani et al 2015). Le modifiche introducono complessità all'interno di un corpus normativo in termini di certezza del diritto, chiarezza del testo aggiornato, semplicità di applicazione delle norme. Tuttavia sono necessarie modifiche, soprattutto in nuovo dominio con una forte emergenza, per riferimenti a casi particolari, per specificare il dominio di applicazione delle norme, per correggere errori materiali e sostanziali, per estendere i termini delle scadenze procedurali. L'analisi delle modifiche permette di valutare l'efficacia delle azioni normative in situazioni di emergenza e consente di fornire istruzioni al legislatore per migliorare l'efficacia delle norme. L'analisi svolta da Palmirani et al (2015 e 2016) sul corpus di ordinanze emesse nei primi 18 mesi dopo il sisma in Emilia, sono state rilevate 820 modifiche su 90 ordinanze (dal 22 giugno 2012, il 17 ottobre 2013); 80% entro il 2012 (7 mesi dall'evento). Il lavoro di identificazione delle modifiche è stato condotto manualmente da esperti legali che hanno costruito il ciclo di vita dei documenti normativi in esame.

Un terzo filone di ricerca riguarda i controlli di robustezza delle metodologie adottate. Un primo confronto riguarderà la categorizzazione dei temi condotta a partire dall'analisi automatica e dalla analisi delle reti di citazioni condotta con Infomap, con l'analisi soggettiva condotta da giuristi esperti che hanno codificato le stesse le ordinanze in modo manuale, utilizzando il vocabolario di tematiche e sotto categorie identificato dall'analisi cluster riportata nella sezione 6.1. Un secondo confronto riguarderà il complesso di citazioni di atti normativi identificabili con le queries specificate nella sezione 5 con la rilevazione automatica dei riferimenti normativi all'interno dei testi realizzata utilizzando *parser* basato su espressioni regolari, modelli e vocabolario giuridico (Palmirani, Cervone, et. al., 2013). Dal punto di vista metodologico questo secondo metodo offre un controllo per la robustezza del cluster analysis prodotta attraverso l'analisi automatica del testo presentata in questo lavoro e per l'analisi di rete che è oggetto del lavoro in corso di Pavone, Righi e Russo (2016).

Ci aspettiamo che nel complesso, oltre alle implicazioni di tipo metodologico sopra richiamate, tali analisi possano produrre alcune raccomandazioni per migliorare la legislazione nazionale sull'emergenza, attualmente in discussione al Senato, che non sarebbe stato possibile definire in modo rigoroso con la lettura dei singoli ordinamenti e della legislazione.

In particolare, il nesso tra emergenza e ricostruzione è una chiave di lettura della produzione normativa presa in esame e si intende mettere in luce come l'azione del Commissario Delegato alla ricostruzione in Emilia-Romagna possa costituire un modello di riferimento per la risposta a disastri naturali: un modello caratterizzato non già da una cesura tra emergenza assistita e successiva azione di ricostruzione, ma un intreccio che anticipa quanto più possibile la ricostruzione per poter limitare il più possibile la fase di emergenza.

Bibliografia

- Balbi, S. & Misuraca, M., 2005. Pesi e metriche nell'analisi dei dati testuali. *Quaderni di Statistica*, Issue 7, pp. 55-68.
- Benzecri, J., 1973. L'analyse de données. Vol II L'Analyse des correspondances. Paris: Dunod.
- Benzecri, J., 1981. Pratique de l'Analyse des Données, tome 3: Linguistique et Lexicologie. Paris: Dunod.
- Bolasco, S., 1999. Analisi Multidimensionale dei dati. Metodi, strategie e criteri di interpretazione. (5° rist. 2012) Roma: Carocci.
- Bolasco, S., 2010. TaLTaC 2.10. Sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi. Milano: LED.
- Bolasco, S., 2013. *Analisi Automatica dei Testi*. Roma: Carocci.
- Bolasco, S., Canzonetti, A. & Capo, F., 2005. Text Mining: Uno strumento strategico per imprese e istituzioni. Roma: CISU.
- Bolasco, S. & Pavone, P., 2008. Multi-class categorization based on cluster analysis and TFIDF. In: *JADT 2008. The 9th International Conference on the Statistical Analysis of Textual Data..* Lyon: s.n.
- Bolasco, S. & Pavone, P., 2010. Automatic Dictionary and Rule-Based Systems for Extracting Information from Text.. In: F. PALUMBO & C. N. LAURO, a cura di *Data Analysis and Classification. Proceedings of the 6th Conference of the Classification and Data Analysis Group of the Società Italiana di Statistica*. Berlin: Springer, pp. 189-198.
- De Domenico, M., Lancichinetti, A., Arenas, A. & Rosvall, M., 2015. Identifying Modular Flow on Multilayer Networks Reveals Highly Overlapping Organization in Interconnected Systems, *Physical Review X*, 5, 011027
- Lebart, L. & Morineau, A., 1984. SPAD. In: *Analyse de Données Textuelle*. Paris: Dunod.
- Lebart, L. & Salem, A., 1988. Analyse statistique des données textuelles. Paris: Dunod.
- Lenci, A., Montemagni, S. & Pirrelli, V., 2005. *Testo e Computer. Elementi di linguistica computazionale*. Roma: Carocci.
- Palmirani, P., Bianchi, I., Cervone, L., & Draicchio, F., 2015. Analysis of Legal References in an Emergency Legislative Setting, paper prepared for the JURIX 2015, 28th International Conference on "Legal Knowledge and Information Systems", Braga, Portugal, 9-11 December, 2015
- Palmirani, P., Bianchi, I., Cervone, L., & Draicchio, F., 2016. Normative References Analysis for Enhancing the Legislative Effectiveness in the Emergency Situations. The case of the earthquake in Emilia-Romagna, paper submitted to JADT2016 Journées internationales d'Analyse statistique des Données Textuelles, 7-10 Juin 2016 Nice (France)
- Pavone, P., 2010. Sintagmazione del testo: una scelta per disambiguare la terminologia e ridurre le variabili di un'analisi del contenuto di un corpus. Roma, LED, pp. 131-140.

- Pavone, P., Righi, S., Russo, M., 2016. Text mining and network analysis to support improvements in legislative action. The case of the earthquake in Emilia-Romagna, paper submitted to JADT2016 Journées internationales d'Analyse statistique des Données Textuelles, 7-10 Juin 2016 Nice (France)
- Piazzì, V., Pagliacci, F. & Russo, M., 2015. Analisi cluster delle caratteristiche socio-economiche dei comuni dell'Emilia-Romagna: un confronto tra comuni dentro e fuori dal cratere del sisma, DEMB Working Paper Series, n. 61
- Rosvall, M. & Bergstrom, C.T., 2008. Maps of information flow reveal community structure in complex networks, PNAS 105, 1118
- Salton, G., 1989. Automatic Text Processing: the Transformation, Analysis and Retrieval of Information by Computer. New York: Addison Wesley.
- Sullivan, D., 2001. Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing and Sales. New York: John Wiley & Sons Inc..
- Zadeh, L., 1977. Fuzzy sets and their applications to pattern classification and clustering. In: V. R. J., a cura di *Classification and Clustering*. New York: Academic Press.
- Zanasi, A., 2005. Text Mining and its Applications to Intelligence, CRM and Knowledge Management. Southampton: WITPress.