

Text mining and network analysis to support improvements in legislative action. The case of the earthquake in Emilia-Romagna

Pasquale Pavone¹, Riccardo Righi², Simone Righi², Margherita Russo³

¹ University of Padua and University of Modena and Reggio Emilia, Italy

² University of Bologna and University of Modena and Reggio Emilia, Italy

³ University of Modena and Reggio Emilia, Italy

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 and policy makers 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 the decrees related to the 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: Contribution grant criteria; 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. Furthermore, by adopting an hybrid system of text mining, it is possible to extract the legislation (and other ordinances) cited in the ordinances and identify clusters of regulatory areas of reference to meet the emergency and reconstruction following natural disasters. In this second strand, clusters of citations are detected with an algorithm of network analysis (Infomap) based on information theory (Rosvall & Bergstrom, 2008; De Domenico et al, 2015). This analysis highlights subsets of nodes and allows to outline which are the most relevant issues involved in the areas of intervention after a natural disaster.

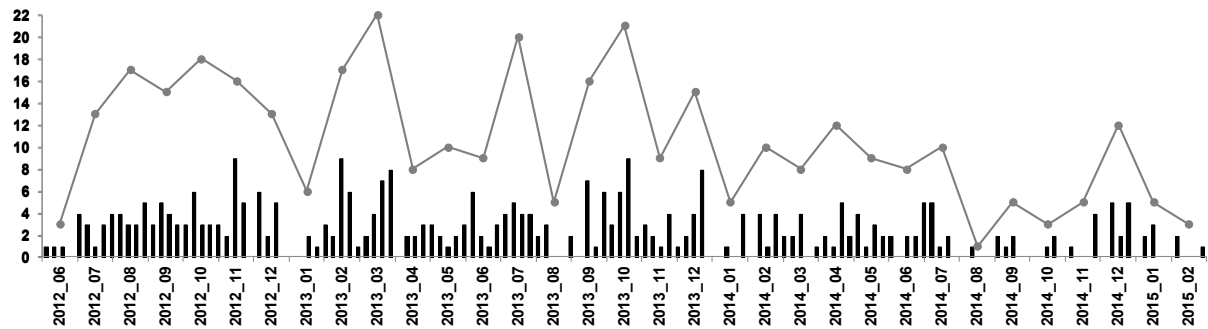
Keywords: automatic classification; legislative corpus; clustering and network analysis.

1. Introduction

In the three years after the 2012 earthquake in Emilia-Romagna, through the enactment of more than 350 ordinances, the Commissioner nominated by the national government 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 and policy makers 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. As a result, a law has been recently passed leaving to further decrees all the details for specific interventions. Within the research project Energie Sisma Emilia, the analysis of the ordinances

was started to study the organizational process put in place to deal with the emergency after an earthquake and to outline the general legal framework to which those ordinances refer. Ordinances issued by the Commissioner, considered in our analysis, cover a period running from June 2012 to February 2015. All the ordinances were downloaded and a preliminary temporal analysis of the normative production is shown in Figure 1

Figure 1 Number of ordinances of the Commissioner published from June 2012 to February 2015, by week (bars) and month (grey line)



To disentangle the areas of intervention of ordinances issued after the earthquake and their legal framework, we adopt two complementary tools: content-text analysis and multilayer network analysis.

As most normative acts, each ordinance is divided into three text sections: the first part of *Introduction*; a second part of *Provisions*; a third part of *Attachments*. For the analysis, it was decided to consider only the text of the “*Introduction*” and “*Provisions*”, excluding the *Attachments* section. Indeed, the latter contains mainly empty forms and texts recalled from previous ordinances. The resulting corpus is composed of 349 Ordinances promulgated from 08/06/2012 to 02/23/2015. After the first parsing of the text, the Vocabulary is composed of 31.969 original graphic forms, for a total of 1.278.291 occurrences. Built the ordinances’ database, we first identify the issues they address.

In absence of predefined categories, the documents are generally grouped according to their semantic similarity, identifying a posteriori their predominant theme. In most of the text mining procedure, a classification of this type leads to disjunctive classes, even though the themes have semantic elements or characteristics in common. The first methodology leads to univocal classification, the second to fuzzy classification (Zadeh, 1977; Ricolfi, 1992). Using a procedure presented in a previous work (Bolasco and Pavone 2008), we can exploit the first type of disjoint classification to produce the second and, in particular, the goal is to obtain a non-supervised clustering to be used in a multi-class supervised categorization. The sets of terms characterizing each cluster is used to obtain a refining of disjunctive classification towards a fuzzy multi-class categorization. This will be achieved through three phases: Information Extraction (IE); cluster analysis and text categorization (section 2). Noisy ordinances detection is described in section 3.

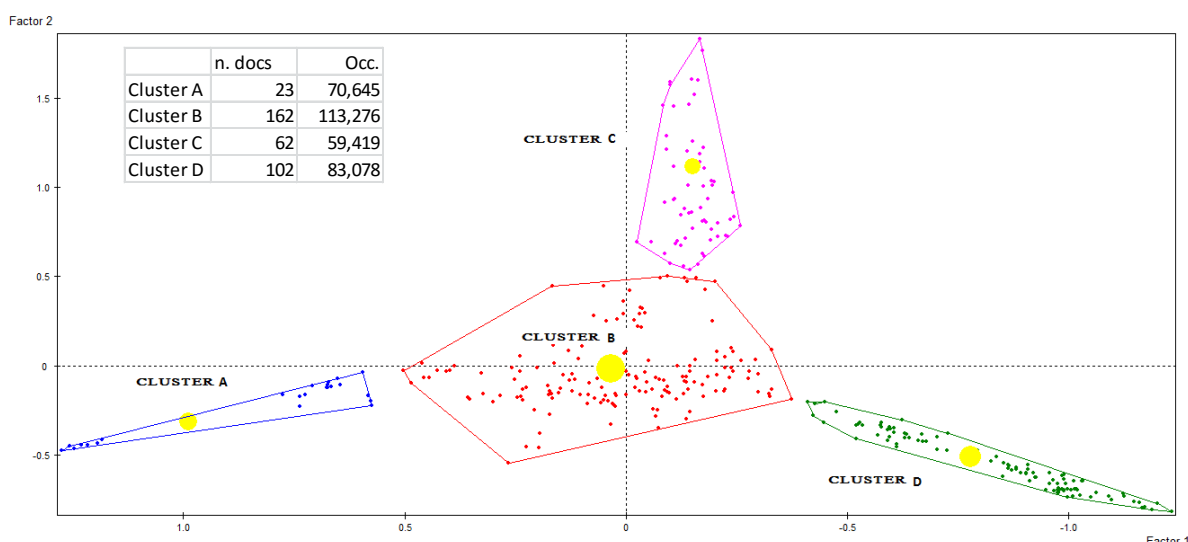
Furthermore, by adopting an hybrid system of text mining, integrating lexical analysis and textual analysis (Bolasco and Pavone, 2010), the automatic extraction of the normative references cited in the documents is possible (section 4). This output allows the identification of clusters of regulatory references to meet the emergency and reconstruction following natural disasters. In this strand, clusters of citations were detected with an algorithm of network analysis (Infomap) based on information theory (Rosvall and Bergstrom, 2008; De Domenico et al, 2015) (section 5). This analysis highlights overlapping subsets of nodes intensely connect-

ed and allows to outline which are the most relevant issues involved in the areas of intervention after a natural disaster. Section 6 concludes discussing some policy implications of our results and outlining some further development in the research.

2. Thematic areas covered by the ordinances

In order to uncover the thematic areas covered by the ordinances, we first proceeded with a lexical analysis, characterized by grammatical tagging and identification of multiword. Then we followed up by identifying the issues of the ordinances through a combination of *Correspondence Analysis* and of *Cluster Analysis* applied on the matrix [*Documents x Keywords*] (349 x 4.273). In particular, the 4.273 keywords were selected in the previous step of lexical analysis, through peculiar language values and grammatical features, selecting only the forms noted as nouns, adjectives and verbs. In Figure 2 are represented the convex hulls of the 4 groups of documents on factorial plane f1-f2.

Figure 2 Convex hulls of the 4 groups of ordinances on factorial plane of the matrix [*Documents x Keywords*]



Observing the distribution of documents and partitions on the factorial plane it is possible to distinguish a central larger barycentric group (Cluster B) from which three separate queues of documents range in, generating three different clusters of ordinances very polarized with respect to each other. The words that characterize each cluster produced a dictionary of terms. The following themes, summarized below, were assigned to the four groups:

Cluster A (23 documents – 70.645 occ.) – **Contribution grant criteria**

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 B (162 documents – 113.276 occ.) – **Management of Resources' Allocation**

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, ter-

ritorio, destinatario, assegnatario, nuclei familiari, Nucleo di Valutazione, Amministrazioni, Comuni, Chiese, edilizia scolastica, alloggi, edifici di culto;

Cluster C (62 documents – 59.419 occ.) – **Urgent Works for Municipalities, Schools and Churches Buildings**

Edifici scolastici, anno scolastico, Scuole, opere di urbanizzazione, riparazione immediata, prefabbricati modulari scolastici, riapertura delle scuole, Edifici Pubblici Temporanei, realizzazione, localizzazione, Programma Operativo Municipi, celerità, urgenza, Prefabbricati Modulari, contributo integrativo, riparazione immediata, Costruzione, appalto, municipi, straordinaria, esito di agibilità, riparate, strutture scolastiche, scuole paritarie, edifici municipali, prefabbricati modulari municipali, edifici religiosi, chiese temporanee, esercizio del culto, collaudo tecnico;

Cluster D (102 documents – 83.078 occ.) – **Interventions to Support Population**

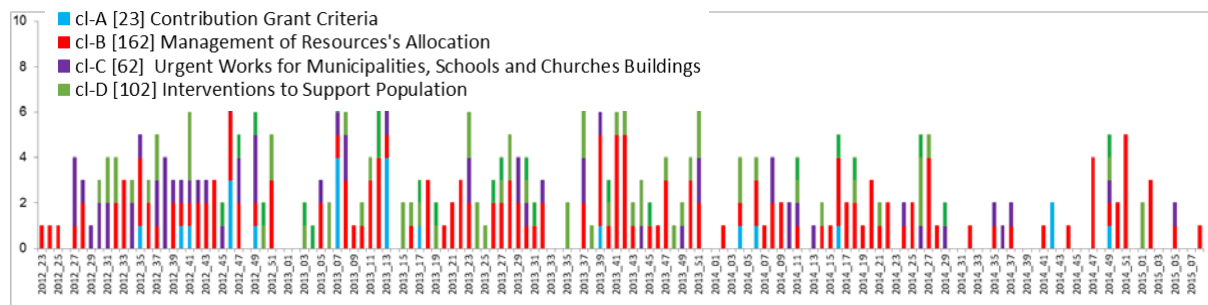
interventi provvisionali, urgente, oneri finanziari, protezione civile, vigili del fuoco, stato di emergenza, amministrazioni locali, oneri di natura sociale, interventi di soccorso, assistenza, popolazione, incolumità, anziani, provvisoriale, disabili NO autosufficienti, accoglienza, contabilità speciale, sanitarie, coordinamento operativo, strutture ricettive, assistenza specialistica, volontariato, salvaguardia, strutture sociosanitarie, attività di assistenza, campi.

From a semantic point of view, it is possible to give an interpretation of the factorial axes thanks to thematic characterization of each group of documents. The factorial axis 1 sharply contrasts “*Contribution grant criteria*” (cl-A) and the “*Interventions to support population*” (cl-D), so the first factor explains the transition from the definition of criteria towards a practical aspect of emergency measures to be taken. Similarly, the polarization on the factor 2 between “*Urgent Works for municipalities, schools and churches buildings*” (cl-C) and “*Interventions to support population*” (cl-D) highlights, within the scope of the urgency of the interventions, the transition between the restorations of fundamental places of public life and the measures provided to direct assistance of the population.

2.1. Temporal analysis of the issues

Figure 3 presents the ordinances shown in Figure 1, classified according to the theme assigned with the procedure discussed in the previous section. This representation highlights the temporal dynamics of the areas of intervention during the reporting period.

Figure 3 Number of ordinances of the Commissioner, by week and cluster (June 2012- February 2015)



The ordinances concerning the “management of resources allocation” [cl-B], the largest group that consists of 162 ordinances, have been the first to be enacted and that continue to be issued throughout the period considered on an almost weekly basis. In the first month there are no ordinances to regulate “interventions to support population” [cl-D], in fact those are typi-

cally under the Civil Protection management; they start only when the Civil Protection ends its specific interventions of assistance. “Urgent works for municipalities, schools and churches buildings” [cl-C] appear the first measures to restore public life and are issued on a regular basis over the next three years: these ordinances structured the transition from emergency to reconstruction and outline the conditions to make possible the reactivation of schools (with school reconstruction plan) and the restoration of the territorial government sites (municipalities) and worship (churches). Finally, punctuating the whole period, there are ordinances on “contribution grant criteria” [cl-A]: they are the smallest group of our field of observation (23 orders).

2.1. Supervised multi-class categorization of the ordinances

The multi-class categorization is based on TITFIDF index¹ (hereafter TI index) (Salton, 1989), for our purpose we consider the four thematic dictionaries resulting from cluster analysis, as textual queries against which to calculate the TI for each document. These TIs are the internal elements of a matrix [*Documents x Queries*] (349 × 4). The value of the TI index, calculated for each thematic lexicon with respect to each ordinance, allows to measure the relevance of a document with respect to the four themes. To obtain the fuzzy belonging value of the *i*-th document with respect to the *j*-th query (topic) we relativize the TI values to the total scores for the document (Pavone and Romano, 2013). The new matrix contains the fuzzy membership values of documents with regard to the four themes. By assuming an equal distribution of the themes across documents, we assign the membership to a theme whenever the degree of membership exceeds the level of equidistribution, which in case of four categories is equal to 0.25.

Table 1 Number of ordinances according to the disjoint categorization and to fuzzy categorization

disjoint categories		multi-class categorization (fuzzy)				
		CAT-1	CAT-2	CAT-3	CAT-4	
A	Contribution grant criteria	23	CAT-1 21	9	0	0
B	Management of Resources' Allocation	162	CAT-2 9	136	15	9
C	Urgent Works for Municipalities, Schools, Churches	62	CAT-3 0	15	60	1
D	Interventions to Support Population	102	CAT-4 0	9	1	97
total		349				

About 10 per cent of ordinances (33 out of 349) have multiple classification, with overlappings mainly insisting on Cluster B (Management of Resources' Allocation). This result is a consequence of the fact that B is the central cluster in the factorial plan, and thus it is (on average) representative of the themes present in the corpus of ordinances. This outcome stresses the robustness of the multi-class categorization adopted in the remaining part of our analysis.

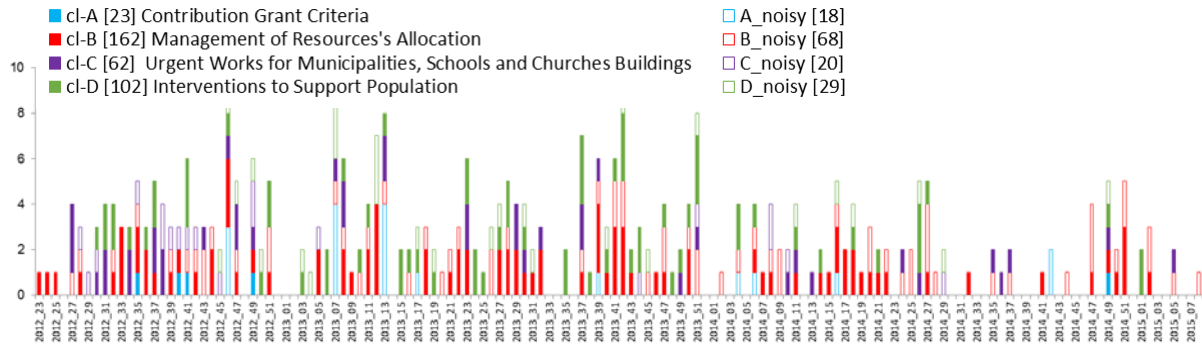
3. Identification of noisy ordinances

Ordinances may contain additions, replacements, adjustments, extensions, changes of previous ordinances. In our data base almost 4 out of ten ordinances (135) contains one or more

¹ Term frequency–inverse document frequency index.

changes of the original ones², in all the categories and along the entire period (as shown in Figure 4). Clusters of ordinances regarding contribution grant criteria and those on the management of resources have the largest share of those changes (respectively 78,3% and 42%). The changes occur during the whole observed time period (see Figure 4). In the following we name this group of ordinances as "noisy ordinances".

Figure 4 Number of ordinances by week and cluster, and by type (June 2012- February 2015)



4. Identification of cited legislations in ordinances

In order to identify the legislation references cited in the ordinances, we have adopted the formalization of their syntactic structure, characterized by the main elements and some optional elements. A legislative act is defined by the sequence of at least two basic elements: the type of act (law, decree, ordinance) and the date of promulgation. Within these two elements it is possible to find the number of the act, or preposition. In some cases, the number of the act may be fused with the date in a single graphic form of type nn/year (e.g. 12/2012). The formalization procedure is characterized by three steps: formalization of all cited types; tagging of types in vocabulary; running the regular expression. The search for legislation is done by defining a Regular Expression in Talend which sequentially makes annotations made on the types (words):

```
"CATSEM (LegAc) CATSEM (da
```

This application individuated 18,715 entities related to 2,298 different graphic forms corresponding to 486 legislative acts (259 ordinances and 227 normative acts) cited in the 349 ordinances of our corpus. Table 2 reports the most cited normative acts recognized through the procedure. The two most frequent occurrences refer to the regulatory norms implemented in 2012, at national level, after the earth quake (the decree and its conversion in law). The third and fourth most cited regulatory acts are the decree and its conversion in law of the national growth plan (framing the resources for the area hit by the earth quake). The next two refer to simplification of procedures and norms on public contracts. The most cited ordinance is the "Ordinanza n. 57 del 12 ottobre 2012", pillar for ordinances in categories A and B;. Large amount of references regards the National Civil Protection and the ordinances issued by it for the hit area in Emilia-Romagna.

² In particular: 67 were additions, 5 replacements; 28 contained adjustments, 12 extensions and 81 ordinances had changes. For a detailed analysis see Palmirani et al. (2015)

TEXT MINING AND NETWORK ANALYSIS

Table 2 - Main legislation and their occurrences, recognized in the two sections analyzed of the data set (349 ordinances)

Legislations	Occ.	cited in ordinances by mult-categories			
		A	B	C	D
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	28	168	74	107
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	29	160	64	101
DECRETO-LEGGE 22 giugno 2012, n. 83. Misure urgenti per la crescita del Paese.	410	8	64	73	4
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	22	69	64	2
LEGGE 24 novembre 2000, n. 340 Disposizioni per la delegificazione di norme e per la semplificazione di procedimenti amministrativi - Legge di semplificazione 1999.	389	20	85	52	39
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	14	40	20	6
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 sostituzione delle scorte e dei prodotti e per la delocalizzazione, in relazione agli eventi sismici del 20 e 29 maggio 2012	291	20	43		
LEGGE 24 febbraio 1992, n. 225. Istituzione del Servizio nazionale della protezione civile.	276	16	93	73	98
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	1	47	73	85
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 Ufficiale n. 130 del 6 giugno 2012	258		3		73
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	20	34		2
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		5		83
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	26	31	1	11
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	10	73	16	55

5. Multilayer analysis of the normative act-ordinances dataset

By defining a specific textual lexicon model, we extract the legislation (and other ordinances) cited in the ordinances to identify clusters of regulatory areas of reference that characterize the emergency and reconstruction following natural disasters. Such clusters might help in designing a new regulatory standard to be referred more easily according the specific need to be addressed during the emergency phase.

In this strand, citations determine links from one ordinance to a normative act or to another ordinance. The resulting network is connected, sparse (density equal to 0.04) and assumes the characteristics of a small world network (clustering about 0.3; average shortest path 2.58). Clusters of citations are detected with Infomap, an algorithm of network analysis (Rosvall *et al.*, 2009) based on information theory (Rosvall and Bergstrom, 2008; De Domenico *et al.*, 2015). This analysis makes possible to outline which are the most relevant issues involved in

the areas of intervention after a natural disaster and the possible overlaps among them³. In this section we briefly present the references to the multilayer model, then we comment the main results.

Multilayer network

By adopting the analysis of multilayer networks (recently developed by De Domenico et al., 2015), we identify the emerging clusters of regulatory acts (ordinances and normative acts) cited by the ordinances. In particular, we adopt a method introduced by Rosvall and Bergstrom (2007 and 2008) that solves the main problems with standard community detections methods (Newman and Girvan, 2004), especially by making possible to identify communities of very different sizes. It operates by minimizing the description length of a network and the loss of information due to the clustering. De Domenico *et al.* (2015) extend the setup to multiplex networks, showing that by taking into account the multilayer structure of networks new features emerge from nodes interacting in the different layers. A random walker is used to compute flows among nodes in the same layers. With some probability (such as $r=0.15$, as in De Domenico *et al.*, 2015) the random walker jumps across layers. The procedure is similar to the teleportation mechanism in the PageRank algorithm. Communities maximize the probability of the random walker to remain within a group of nodes when starting from one of the nodes within that group. If two nodes in two different layers tend to be visited with similar patterns they are associated to the same cluster which becomes a multi-layer cluster. Thus the algorithm is able to identify both communities identified in one single layer and communities identified on multiple layers. As layers are themselves informative, the outcome is a more realistic and informative clustering.

In our analysis, each layer corresponds to one of the four categories of interventions, defined by the categorization described in section 2.1. Since we adopt the multiclass categorization (fuzzy) summarised in Table 1, then an ordinance may belong to more than one layer (it happens with almost 10% of the ordinances). Normative acts are assigned to all categories of nodes citing them, which are potentially more than one. Links are directed and represent citations of a normative act or an ordinance by an ordinance.

Results

Infomap generates 65 non overlapping clusters⁴. In Figure 5, those clusters are ordered according to decreasing size and, for each Infomap cluster, three types of node are highlighted: ordinances, noisy ordinances and normative acts. The 15 largest Infomap clusters contain almost 83% of the nodes, 13 Infomap clusters have from 2 to 9 nodes, and 37 Infomap clusters have only one node (30 are normative acts and 7 are noisy ordinances). The graph of the 65 Infomap clusters (Figure 6) represents the citations (grey edges) from non-noisy ordinances to other ordinances or to normative acts. Each node is proportional to the number of layers (i.e. the categories defined in Table 1) in which it is active⁵; the colour is the one corresponding to

³ Indeed, this method highlights overlapping subsets of nodes.

⁴ Since there was no reason to exclude overlapping clusters, the adoption of Infomap to multilayer analysis was considered more appropriate than other methods that maximize modularity producing disjoint clusters.

⁵ Each normative act may be cited by ordinances classified in one of the four categories, but the largest part (544 acts) is associated to only one category, 31 to two categories and only one to three categories. In this graph the 135 noisy ordinances are not represented.

the cluster associated by Infomap for that node. In the graph we have computed the association of nodes to the four categories of ordinances (triangles), but to simplify the graph these edges are not shown. To summarize the results presented in this graph we present a heat map (Figure 7) and an aggregated graph (Figure 8).

Figure 5 Size of the 65 Infomap clusters, by type of node

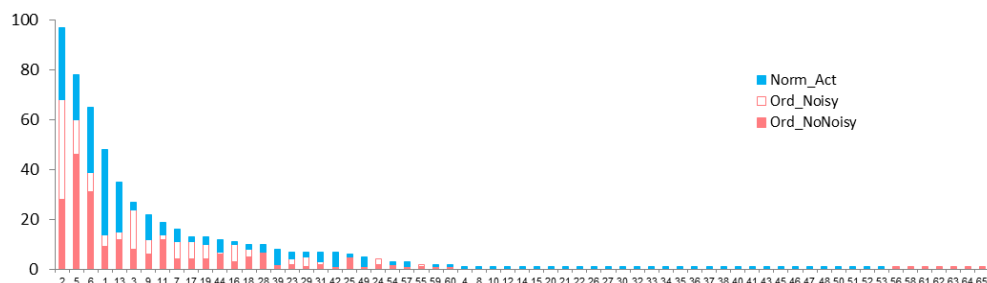
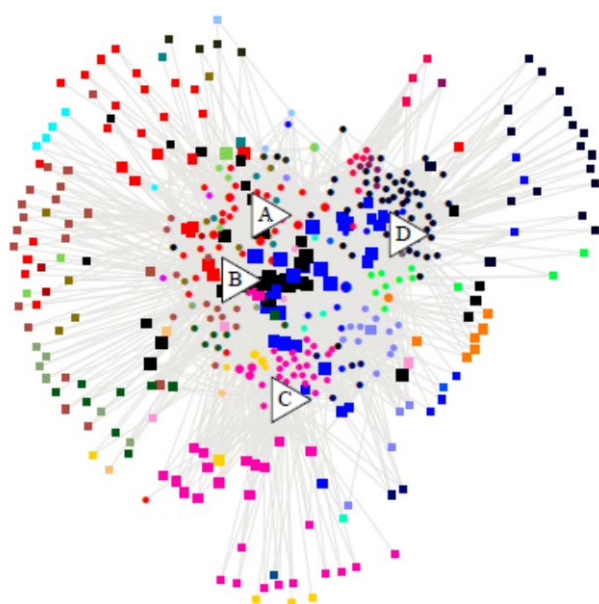


Figure 6 The 65 Infomap-clusters of citations of non-noisy ordinances and normative acts (made in the ordinances), by category of intervention



Vertices

Triangles: the four categories from Table 1 ("A" Contribution grant criteria; "B" Management of Resources' Allocation; "C" Urgent Works for Municipalities, Schools, Churches; "D" Interventions to Support Population).

Circles and squares, respectively: non-noisy ordinances and normative acts cited by the ordinances; size is proportional to the number of layers in which the node is active (min 1, max 4); colours are different by Infomap cluster. Noisy ordinances are omitted.

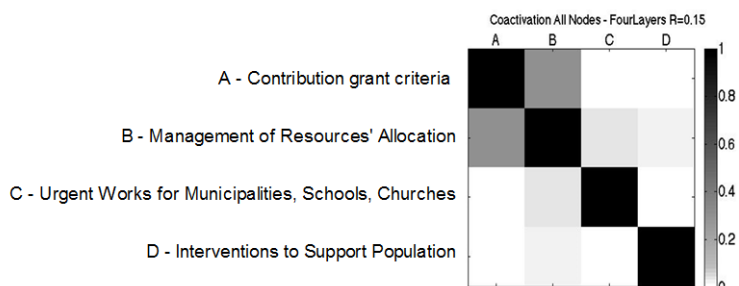
Edges

Greyedges represent citations; number of citations is not weighted.

Edges representing affiliation to Categories of ordinances are computed in the graph but not shown.

R-igraph package Fruchterman Reingold Layout Nodes

Figure 7 Heat map: fraction of nodes in different layers (categories in Table 1) that are assigned to the same Infomap cluster (see Figure 6)



The heat map in Figure 7 highlights the fraction of nodes in different layers (categories in Table 1) that are assigned to the same Infomap cluster: we observe that regulatory references (ordinances and normative acts) largely overlap across interventions on Contribution grant criteria and Management of Resources' Allocation; the latter overlaps also with issues on Ur-

gent Works for Municipalities, Schools and Churches Buildings. Interventions to Support Population appear to have a quite distinct corpus of legal references.

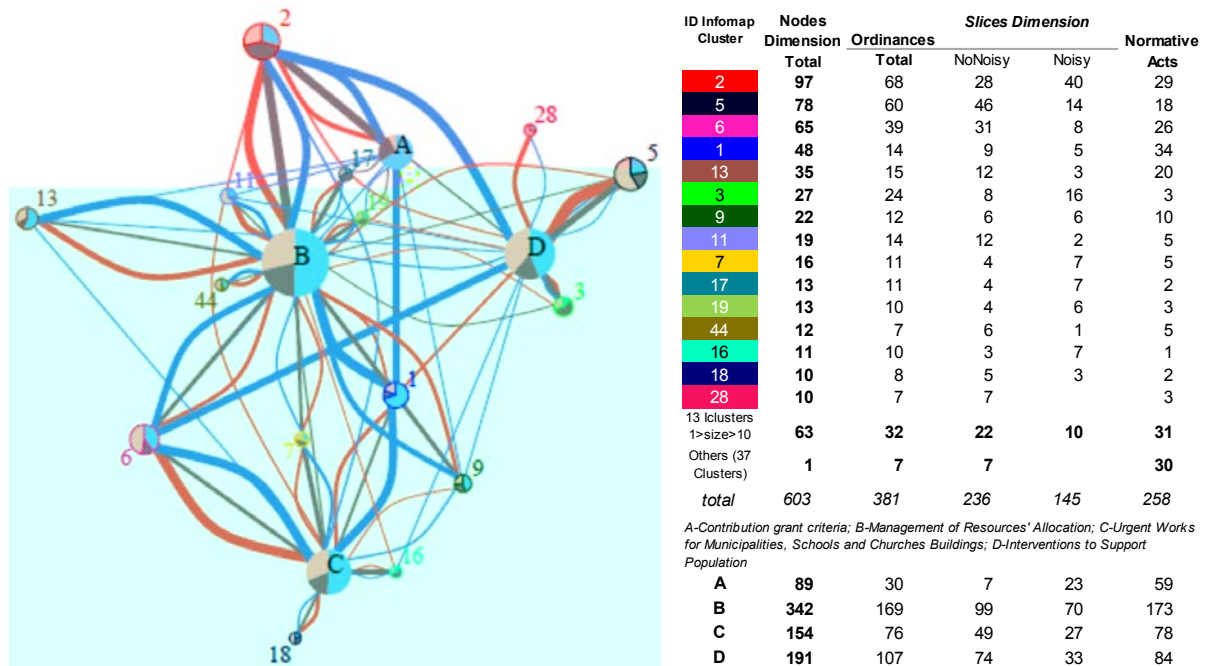
These results partly descend from the multiclass categorization of ordinances (section 2.1) but Infomap clustering reveals the specific subgroups of documents and relative contents involved in those overlapping across categories. A more detailed view on that regulatory environment of references may be explored with the graph in Figure 8, providing a coarse-grained representation of the 15 largest Infomap clusters from the graph in Figure 6.

The largest Infomap cluster (id 2), with 97 nodes, has the largest share of noisy ordinances (40). Nodes belonging to this cluster highlight the main overlapping with the two categories of interventions - contribution grant criteria and managing of resources - observed above, although some nodes are also categorized in the intervention categories (public works and assistance of population). Cluster 5 (78 nodes) involves mainly ordinances and normative acts on urgent works (category D) and some references to managing of resources. Cluster 6 (65 nodes) contains ordinances mainly for urgent works, and includes the basic normative acts cited also in categories B and D. Cluster 1 (48 nodes) contains the largest share of normative acts (34) with ordinances in the area overlapping across categories B, C and D, has references to the same nucleus of normative acts shared by ordinances in categories A. For the remaining clusters in this top list, the overlapping categories are not relevant and in three clusters are absent (id 44, id 18, id 28).

Figure 8 The 15 largest Infomap clusters of ordinances and normative acts (size ≥ 10) and the four categories of interventions (A, B, C, D) they belong to

Vertices: circles with alphabetic letter (A, B, C, D) are the Categories of intervention (see section 2); circles with numbers are the 15 largest Infomap Cluster. Colours of borders (vertex frames) and labels are the same of Infomap clusters in Figure 6, listed in the table on the side. Size of nodes is proportional to the number of ordinances and normative acts that belong to that node. Slices of vertices are proportional to the number of non-noisy ordinances (light red), noisy ordinances (gray) and normative acts (blue) belonging to that node.

Edges represent, respectively, non-noisy ordinances (red edges), noisy ordinances (greyedges) and normative acts (blue edges) belonging to the Infomap Cluster and to the Taltac Category. Width of edges is proportional to the number of ordinances/normative acts belonging to the two involved nodes.



6. Conclusions and further research

Network analysis combined to text analysis supports some policy recommendations that could not have been reached by reading the individual ordinances and legislation. From our analysis of the temporal issues associated with the ordinances, there is a clear implication for the decrees that should be issued to make the national law effective. In this respect we show that Emilia-Romagna's specific set of ordinances is a model for after-earthquake interventions: actions anticipate as much as possible the reconstruction to limit as much as possible the emergency phase. A conclusion that contrasts the current view that it should be defined by law a duration benchmark for emergency assistance after which a subsequent reconstruction effort should be started. A critical aspect, as observed also by Palmirani *et al.* (2015), is that - due to the absence of a clear regulatory frame - almost 40% of ordinances were revisions, substitutions or integrations. This situation has caused slowdowns and delays in effective interventions.

Clustering produced by Infomap makes possible to highlight the structure of the national regulatory environment related to the main areas of interventions and the specific series of ordinances cited in the entire corpus. This issue has been so far analysed by expert analysis (Palmirani *et al.*, 2015). A comparison of these results with the automatic analysis proposed in this work would provide useful insights for both methods.

Moreover, clustering produced with automatic text analysis has also provided an input for a more detailed tagging of each ordinance, according to the categories and sub topics detected in our analysis. This tagging will be applied to compare the cluster analysis produced through automatic text analysis, and clusters detected by Infomap.

References

- Bolasco S. and Pavone P. (2008). Multi-class categorization based on cluster analysis and TFIDF. *JADT2008*, 1, 209-218.
- Bolasco S. and Pavone P. (2010). Automatic dictionary-and rule-based systems for extracting information from text. In *Data Analysis and Classification* (189-198). Springer Berlin Heidelberg.
- De Domenico M. *et al.* (2015). Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*, 5(1), 011027.
- Palmirani P. *et al.* (2015, December). Analysis of Legal References in an Emergency Legislative Setting, paper prepared for the JURIX 2015. In *28th International Conference on "Legal Knowledge and Information Systems"*, Braga, Portugal.
- Pavone P. and Romano M.F. (2013). Models of Text Mining to measure improvements to doctoral courses suggested by "stella" phd survey respondents. In *Statistica, anno LXXIII, n. 4*.
- Ricolfi L. (1992). Sul rapporto di identificazione: l'interpretazione semantica e l'interpretazione sintattica. *Sociologia e ricerca sociale*.
- Rosvall M. *et al.* (2009). The map equation. *The European Physical Journal Special Topics*, 178(1), 13-23.
- Rosvall M. and Bergstrom C. T. (2008). Maps of information flow reveal community structure in complex networks. In *In Proceedings of the National Academy of Sciences USA*, 105, 1118.
- Salton G. (1989). Automatic text processing: The transformation, analysis, and retrieval of *Reading: Addison-Wesley*.
- Zadeh L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 1(1), 3-28.