

L'utilizzo di tecniche di text mining nell'analisi esplorativa delle sentenze di illegittimità costituzionale. Il caso della Corte costituzionale italiana.

Rocco Mazza¹, Luigi Rullo²

¹University of Naples "Federico II" – rocco.mazza@unina.it

²University of Naples "Federico II" – luigi.rullo@unina.it

Abstract 1

In this paper, we propose the application of text mining methods to decisions of the Italian Constitutional court. The innovative contribution is the use of quantitative analysis method in the area of relations between courts and politics. The aim of this paper is exploring the judicial corpus in order to identify the principal topics thanks to the use of a model for the extraction of latent topics. In particular, we will focus on judicial declarations of unconstitutionality from 2008 to 2018. Our approach consists of the following work phases: (1) construction of the corpus by extracting the judgements from the public database of the Constitutional Court; (2) pre-treatment of the corpus and construction of the *document-term* matrix; (3) extraction of the topics through a specific model; (4) applying labels to topics and identifying issues; (5) study of the topics in relation to the variables that characterize the sentences involved in the analysis. The starting point is the Latent Dirichlet Allocation model. LDA allows to infer the latent structure of topics through by recreating the documents in the corpus considering the relative weight of the topic in the document and the word in the topic, in an iterative way. The exercise of judicial review gives Constitutional Court strong powers. For these reasons, an effective thermometer to understand the expansion of the Court in the political arena is the analysis of judicial declarations of unconstitutionality. As a matter of fact, it represents one of the principal demonstrations of judicial activism, which has been particularly high over the last decade.

Keywords: Content Analysis, Constitutional Court, Italy, Juridical Texts, LDA.

Abstract 2

In quest'articolo proponiamo l'applicazione di metodi di text mining a sentenze delle Corte Costituzionale Italiana. L'apporto innovativo del presente lavoro consiste nell'applicazione di un metodo per l'analisi quantitativa del contenuto a testi prodotti nell'ambito riferibile all'area delle relazioni tra corti e politica. L'obiettivo di quest'articolo è esplorare il corpus giuridico preso in esame al fine di individuare le principali tematiche emergenti tramite l'utilizzo di un modello per l'estrazione di topic latenti. In particolare, il nostro lavoro si concentra sulle sentenze di illegittimità costituzionale dal 2008 al 2018. Il nostro approccio si costituisce nelle seguenti fasi di lavoro: (1) costruzione del corpus mediante estrazione delle sentenze dal database pubblico della Corte Costituzionale; (2) pre-trattamento del corpus e costruzione della matrice documenti per testi; (3) estrazione dei topic mediante uno specifico modello; (4) applicazione di etichette ai topic e individuazione delle tematiche; (5) studio dei topic in relazione alle variabili che caratterizzano le sentenze coinvolte nell'analisi. Il punto di partenza è il modello Latent Dirichlet Allocation Il modello LDA permette di inferire la struttura latente dei topic ricreando i documenti nel corpus considerando il peso relativo del topic nel documento e della parola nel topic, mediante molteplici iterazioni. L'esercizio del sindacato di costituzionalità da parte delle Corti costituzionale dota quest'ultima di forti poteri. La possibilità di dichiarare una norma incostituzionale fa sì che quest'ultime entrino a pieno titolo nei circuiti di policy making. Per queste ragioni, le sentenze di illegittimità costituzionale sono considerate un termometro efficace per comprendere l'espansione della Corte nell'arena politica. Di fatto, esso rappresenta una delle principali manifestazioni dell'attivismo giudiziario, che è stato particolarmente alto durante gli ultimi dieci anni.

Parole chiave : Content Analysis, Corte Costituzionale, Italia, Testi Giuridici, LDA.

1. Introduzione

Nel contributo viene proposto l'applicazione di un modello per l'individuazione di topic latenti a sentenze emesse dalla Corte costituzionale italiana. L'apporto innovativo del presente lavoro consiste nell'applicazione di un metodo per l'analisi quantitativa del contenuto a testi prodotti nell'ambito riferibile all'area delle relazioni tra corti e politica in Italia. L'obiettivo del paper è dunque esplorare il corpus giuridico preso in esame al fine di far emergere le tematiche latenti riuscendo a isolare gli elementi semantici procedurali e ridondanti. In particolare, il nostro lavoro si concentra sulle sentenze di illegittimità costituzionale dal 2008 al 2018. Il nostro approccio si costituisce nelle seguenti fasi di lavoro: (1) costruzione del corpus mediante estrazione delle sentenze dal database pubblico della Corte costituzionale, in questa fase sono state definite anche un set di variabili caratterizzanti i testi; (2) pretrattamento del corpus e costruzione della matrice documenti per testi; (3) estrazione dei topic mediante uno specifico modello; (4) applicazione di etichette ai topic e individuazione delle tematiche; (5) studio dei topic in relazione alle variabili che caratterizzano le sentenze coinvolte nell'analisi. Il punto di partenza è il modello Latent Dirichlet Allocation. La LDA è un modello generativo e probabilistico. Viene applicata per l'estrazione di topic latenti, individuati all'interno di una collezione di documenti. Alla sua base è possibile rintracciare le presenti assunzioni: a) i documenti sono rappresentati come mix di topic, dove a ogni topic è associata una distribuzione di parole; b) i topic sono parzialmente nascosti, latenti per la precisione, all'interno della struttura dei documenti. Il modello permette di inferire la struttura latente dei topic ricreando i documenti nel corpus considerando il peso relativo del topic nel documento e della parola nel topic, mediante molteplici iterazioni. Per quanto riguarda il caso studio scelto, bisogna osservare che l'esercizio del sindacato di costituzionalità da parte della Corte costituzionale dota quest'ultima di forti poteri. La possibilità di dichiarare una norma incostituzionale fa in modo che quest'ultima possa entrare a pieno titolo nei circuiti di policy making. I provvedimenti adottati dai giudici, infatti, presentano ampi margini di discrezionalità che comportano l'inevitabile trasformazione e creazione della legge. Per queste ragioni, le sentenze di illegittimità costituzionale sono considerate un termometro efficace per comprendere l'attivismo giudiziario, che è stato particolarmente evidente durante l'ultima decade. Questo fenomeno è apparso limpidamente nell'ultimo decennio, dato che la Corte costituzionale ha modellato e influenzato il processo di policy making in aree di policy molto salienti, comportando un progressivo rafforzamento rispetto alle altre istituzioni.

2. Nota metodologica

La procedura metodologica si compone delle seguenti fasi: 1. Collezione dei testi e pretrattamento; 2. Applicazione del modello per l'estrazione dei topic; 3. Analisi del contenuto. I testi sono stati estratti dal database ufficiale della Corte costituzionale tramite un accesso diretto. Ogni testo estratto è accompagnato da due variabili (in Tabella 1) che descrivono alcune caratteristiche riferibili all'ambito giuridico, in particolare all'anno di emissione della sentenza e alla tipologia di giudizio emesso dalla corte. Le motivazioni che hanno spinto i ricercatori a raccogliere le presenti variabili con i testi sono la necessità di studiare la distribuzione dei topic in una prospettiva diacronica.

Tabella 1 Variabili associate ai testi

| Variabile | Modalità | | | | | | | | | | | |
|---------------------|-------------|------------|------|------|------|------|------|------|------|------|------|--|
| Anno di riferimento | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | |
| Tipo giudizio | Incidentale | Principale | | | | | | | | | | |

A partire dal corpus così costituito è stato eseguito il pretrattamento dei testi seguendo le seguenti fasi: 1. Normalizzazione del testo; 2. Eliminazione delle parole vuote (stopwords); 3. Eliminazione delle abbreviazioni e delle notazioni giuridiche; 4. Estensione di espressioni abbreviate riferibili a normative, leggi, codici e regolamenti. I testi che compongono il corpus sono atti ufficiali digitalizzati, presentano quindi dei caratteri speciali e dei refusi dovuti a tale processo. Durante la fase di normalizzazione dal testo sono stati rimossi questi caratteri attraverso una codifica in un encoding unico, sono stati eliminati segni di punteggiatura, doppi spazi e tutti i caratteri sono stati trasformati in minuscolo. Per le stopwords è stato utilizzato il dizionario fornito dalla libreria *snowball*¹. Le sentenze della Corte presentano un elevato numero di riferimenti normativi, abbreviazioni e formule tipiche del linguaggio specialistico giuridico. Sono state eliminate le forme strettamente procedurali, mentre i riferimenti ai codici maggiori o a specifiche leggi in linea con l'obiettivo del lavoro sono stati resi nella loro forma estesa (es.: c.c. = codice civile). Alla fine, è stata generata la matrice documenti per termini dalla quale sono state rimosse le parole che presentavano una frequenza inferiore a 3. A partire da questa matrice è stato applicato un modello per individuare i topic latenti e ridurre l'informazione in modo da poter individuare i principali topic di discussione. Successivamente è stata effettuata un'interpretazione basata sulla lettura critica dei risultati del modello isolando i venti termini con probabilità più elevata per ogni topic. Tale lettura è stata condotta utilizzando i riferimenti concettuali individuati nella nostra letteratura di riferimento. Attraverso la riduzione dell'informazione contenuta nel corpus e l'individuazione dei topic latenti è stato possibile esplorare e categorizzare il contenuto dei testi in esame. In conclusione, ogni documento è stato associato ad un topic ed è stato possibile studiare gli incroci con le variabili create durante la raccolta dati.

2.1. LDA topic modeling

Da un punto di vista strettamente metodologico, l'obiettivo del presente lavoro consiste nell'individuare dimensioni semantiche latenti. Per ottenere questo risultato è possibile utilizzare una moltitudine di metodi. Per l'analisi dei dati testuali qui presentata è stato selezionato un *Latent Topic Modeling* (da ora LTM), in particolare una *Latent Dirichlet Allocation* (da ora LDA). Considerata la numerosità dei testi all'interno del corpus e la loro lunghezza abbiamo fatto ricorso ad una tecnica non supervisionata che effettuasse una codifica automatica dei testi basata sull'individuazione di topic latenti. Considerando la vocazione interdisciplinare del lavoro, il modello selezionato permette di restituire una nuova organizzazione delle informazioni in forma compatta e chiara anche per chi non possiede familiarità con la tecnica. Un *latent topic model* è un modello statistico che permette di identificare una serie di temi latenti all'interno di una collezione di testi; la letteratura di

¹ Si veda : <https://cran.r-project.org/web/packages/SnowballC/index.html>

riferimento è concorde sul definire questi come modelli probabilistici che analizzano la somiglianza tra la distribuzione dei termini all'interno del documento con quella di un topic latente (e.g., Blei, Ng, e Jordan, 2003; Griffiths e Steyvers, 2002; 2003; 2004; Hofmann, 1999; 2001). È possibile riassumere questa idea di base in tre punti (Blei, 2009):

1. Rivelare pattern semantici latenti all'interno della collezione di testi
2. Associare i testi a questi pattern, i quali altro non sono che i topic
3. Utilizzare questa annotazione per organizzare, riassumere concettualmente e ricercare questi testi

Questi modelli hanno l'obiettivo di rintracciare la struttura semantica latente all'interno di un corpus partendo da un assunto fondamentale: un documento è considerato come un insieme di argomenti, ognuno dei quali è caratterizzato da una particolare distribuzione di termini e parole (Steyvers e Griffiths, 2007). A partire da questa definizione è possibile definire questi modelli come generativi di nuovi documenti: per generare un nuovo testo viene estratto un topic e successivamente una parola dalla distribuzione corrispondente; il processo deve essere iterato per l'intera lunghezza del documento. Ovviamente il processo può essere invertito, al fine di inferire il documento a partire dai topic a disposizione. Un gran numero di topic models probabilistici vengono utilizzati per la *content analysis* e *information retrieval* (Blei et al., 2003; Griffiths e Steyvers, 2002; 2003; 2004; Hofmann, 1999; 2001), uno di questi è la LDA. Il riferimento più importante al modello è il lavoro di Blei et al. (2003), inoltre il modello è stato ampiamente studiato in Griffiths e Steyvers (2004), Heinrich (2005), Blei e Lafferty (2009), Berry e Kogan (2010); in questa sezione c'è una panoramica del topic model LDA, ampiamente basata sul lavoro degli autori originali. Alla base dell'LDA troviamo l'assunto precedentemente definito, è un modello generativo e inferenziale; noi osserviamo solo documenti e parole, nessun topic, questi sono nascosti, latenti per la precisione, all'interno della struttura dei documenti. L'obiettivo è inferire le distribuzioni di tali topic. Il modello ricrea i documenti nel corpus considerando il peso relativo dell'argomento nel documento e la parola nel topic, in modo iterativo. La LDA è un modello bayesiano e presuppone che le distribuzioni possano essere descritte da una distribuzione di Dirichlet (Blei et al., 2003). Come distribuzione coniugata della Multinomiale, è conveniente utilizzarle come *a priori* e ne consegue la distribuzione può essere uno strumento eccellente per quanto riguarda i problemi di inferenza del modello (Neapolitan, 2003; Balakrishnan e Nevzorov, 2003). Senza dilungarci ulteriormente sulla distribuzione di Dirichlet, possiamo considerarla come la semplificazione multivariata della distribuzione Beta utilizzata nelle statistiche bayesiane (Neapolitan, 2003). Prendendo una funzione a posteriori dalla distribuzione di Dirichlet, parametrizzata da pesi maggiori su un singolo topic, possiamo ricavare una distribuzione che va a comporre ogni mix di topic. Per definire il procedimento schematicamente: da una distribuzione di Dirichlet $Dir(\alpha)$ noi effettuiamo un campionamento casuale che rappresenta la distribuzione dei topic di un documento. Questa distribuzione di topic è θ . Da θ , noi selezioniamo un particolare topic Z basato sulla distribuzione. Successivamente da un'altra distribuzione di Dirichlet $Dir(\beta)$, noi selezioniamo un campione casuale che rappresenti la distribuzione di parole del topic Z , questa distribuzione è ϕ , da ϕ noi scegliamo la parola w .

2.2. Definizione del numero di topic

Prima di procedere con l'applicazione del modello sulla matrice documenti per termini, ottenuta successivamente al pretrattamento, è stato definito il numero di topic individuando il miglior modello. Di solito nei topic model il numero di topic da estrarre è un parametro definito a priori, questo potrebbe non coincidere con le esigenze di un approccio esplorativo in cui si ha intenzione di estrarre automaticamente i topic dalla collezione di testi. Esistono in letteratura una moltitudine di metodi che permettono di valutare il modello con il numero ottimale di topics (Blei and Lafferty, 2009, Wallach et al., 2009; Buntine, 2009; Chang and Blei, 2009). Per far ciò è possibile avviare il modello uno specifico numero di volte e scegliere tra tutte l'iterazione che ottimizza una serie di parametri precedentemente definiti, in pratica la miglior performance: "Estimating the probability of held-out documents provides a clear, interpretable metric for evaluating the performance of topic-related models related to other topic-based models as well as other non-topic-based generative models" (Wallach et al., 2009). In questo paper il miglior modello è stato selezionato tramite il metodo della media armonica (Griffiths et al., 2005; Zheng et al., 2006; Wallach, 2006, Wallach et al., 2009; Buntine, 2009). Di seguito la formalizzazione (Griffiths and Steyvers, 2004): «In our case, the data are the words in the corpus, w , and the model is specified by the number of topics, K , so we wish to compute the likelihood $p(w|K)$. The complication is that this requires summing over all possible assignments of words to topics z [, i.e., $p(w|K) = \int p(w|z,K)p(z)dz$]. However, we can approximate $p(w|K)$ by taking the harmonic mean of a set of values of $p(w|z,K)$ when z is sampled from the posterior $p(z|w,K)$ (Kass and Raftery, 1995)».

3. Il caso studio: L'espansione della Corte costituzionale nell'arena politica

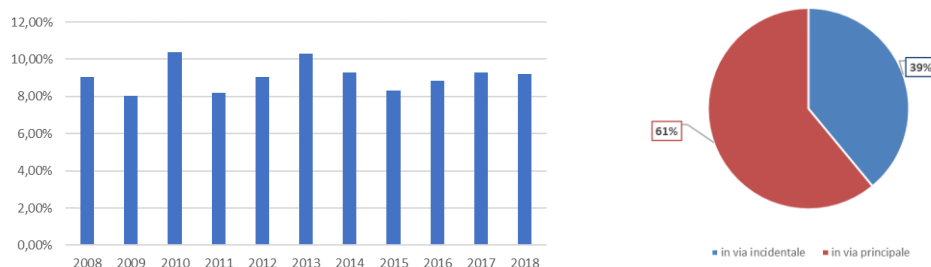
L'espansione del potere giudiziario è divenuto un fenomeno globale negli ultimi decenni (Tate e Vallinder, 1995; Guarnieri e Pederzoli, 2017; Rullo, 2020). In numerosi regimi politici contemporanei si è osservato l'allargamento "del raggio d'azione dei tribunali o dei giudici a scapito dei legislatori e/o amministratori, ovvero lo spostamento delle competenze decisionali dal legislativo, dall'esecutivo o dall'amministrazione verso i tribunali" (Tate e Vallinder 1995, 13). Si pensi alle decisioni della Corte costituzionale spagnola in relazione all'indipendenza della Catalogna, oppure al recente intervento della Corte Suprema inglese sulla sospensione dell'attività del Parlamento. In sostanza, si sta osservando la "dipendenza dalle corti e i mezzi giudiziari per indirizzare questioni morali difficili, questioni di politiche pubbliche, e controversie politiche" (Hirschl 2008,119). Di fatto, questo fenomeno segnala dei mutamenti profondi dei pilastri dei regimi politici contemporanei, ed un graduale ribilanciamento nelle relazioni tra le diverse istituzioni. Si tratta di un processo ancora più evidente quando parlamenti e governi entrano in periodi di elevata frammentazione e scarso rendimento. Il caso italiano è risultato pienamente in linea con queste dinamiche. A partire dagli anni Novanta, l'Italia si è trasformata da una delle più consolidate partitocrazie nello scenario internazionale ad uno degli esempi più chiari di politica personale (Musella, 2018). Lo scandalo di corruzione Tangentopoli ha mandato letteralmente in crisi il sistema dei partiti portando nel 1994 alla nascita della cosiddetta Seconda repubblica. Si è osservato, quindi, uno scivolamento di importanti prerogative al di fuori del tradizionale circuito di rappresentanza, che ha favorito un progressivo rafforzamento del potere giudiziario e in particolare della Corte costituzionale. In questo scenario, le decisioni adottate dai giudici hanno condotto a maggiori frizioni tra giustizia e politica. Basti considerare che la Corte costituzionale ha dichiarato illegittime due leggi elettorali, prima con la sentenza 1/2014, e poi con la decisione 35/ 2017, intervenendo su aree solitamente riservate al legislatore. Per comprendere l'influenza che la Corte costituzionale riesce ad esercitare sull'arena politica bisogna

considerare le dichiarazioni di illegittimità costituzionale (Pederzoli, 2008). Esse rappresentano un indicatore comunemente accettato dalla letteratura internazionale per misurare l'attivismo delle corti e comprendere la natura delle relazioni che intercorrono tra *law* e *politics* (Hirschl, 2008). Ormai la produzione di norme ha luogo inesorabilmente all'ombra della revisione di costituzionalità, e le aree di intervento su cui i giudici sono chiamati ad intervenire divengono sempre più delicate. In questo scenario, comprendere i topic delle dichiarazioni di illegittimità della Corte costituzionale permette di comprendere meglio il suo ruolo all'interno dello scacchiere istituzionale e politico. Per queste ragioni, riteniamo che l'utilizzo di tecniche di text-mining possa rappresentare un primo passo per esplorare le relazioni che intercorrono tra corti e politica in Italia da una nuova prospettiva. Di fatto, lo sviluppo di metodi quantitativi e la possibilità di analizzare un corpus giuridico di sentenze liberamente accessibile dischiudono aree di ricerca in continua evoluzione (Lenci et al. 2009; Montemagni et al. 2010; Sigismondi et al. 2011; Sagri et al. 2014; Dyevre et al. 2019). Come notano Sigismondi et al. (2011, 25) “le tecniche di text mining rendono oggi possibile l'analisi di questi documenti e la realizzazione di strumenti automatici per strutturare, organizzare, esplorare e facilitare l'interpretazione del contenuto di grandi collezioni di documenti testuali”. La prospettiva che guida il lavoro è di natura interdisciplinare, e sottolinea la necessaria collaborazione tra scienze sociali e statistiche per comprendere la realtà empirica.

4. Risultati

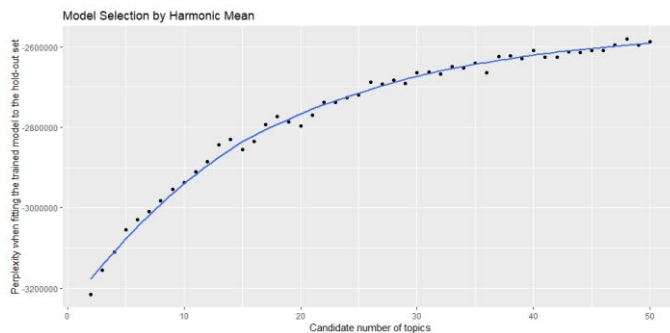
L'analisi prende in considerazione la totalità di sentenze di illegittimità pronunciate dalla Corte costituzionale tra il 2008 e il 2018, il numero complessivo è pari a 1000. Le sentenze sono state scaricate dal sito ufficiale della Corte, ed hanno una lunghezza media piuttosto elevata, pari a 25.038 caratteri. Gli anni in cui sono state prodotte più sentenze sono il 2010 e il 2013, mentre la tipologia di giudizio più comune è quella in via principale. Quest'ultimo si riferisce al caso in cui il giudizio di costituzionalità è promosso dallo Stato contro una legge regionale o da una Regione contro una legge statale. Nel giudizio in via incidentale, invece, il giudizio di costituzionalità è promosso da un giudice nel corso di un processo nel caso in cui riscontri dubbi sulla legittimità costituzionale di una norma di legge di cui deve fare applicazione. Essi rappresentano i principali tipi di giudizio emanati dalla Corte costituzionale.

Fig. 1 a. Distribuzione della variabile anno b. Distribuzione del tipo di sentenza



In figura 2 sono presenti i risultati restituiti dal calcolo del miglior modello. Il calcolo è stato effettuato avviando il modello per 50 volte su una sottomatrice composta dal 50% delle righe e delle colonne estratte casualmente. Il modello con la miglior performance è quello con 48 topic.

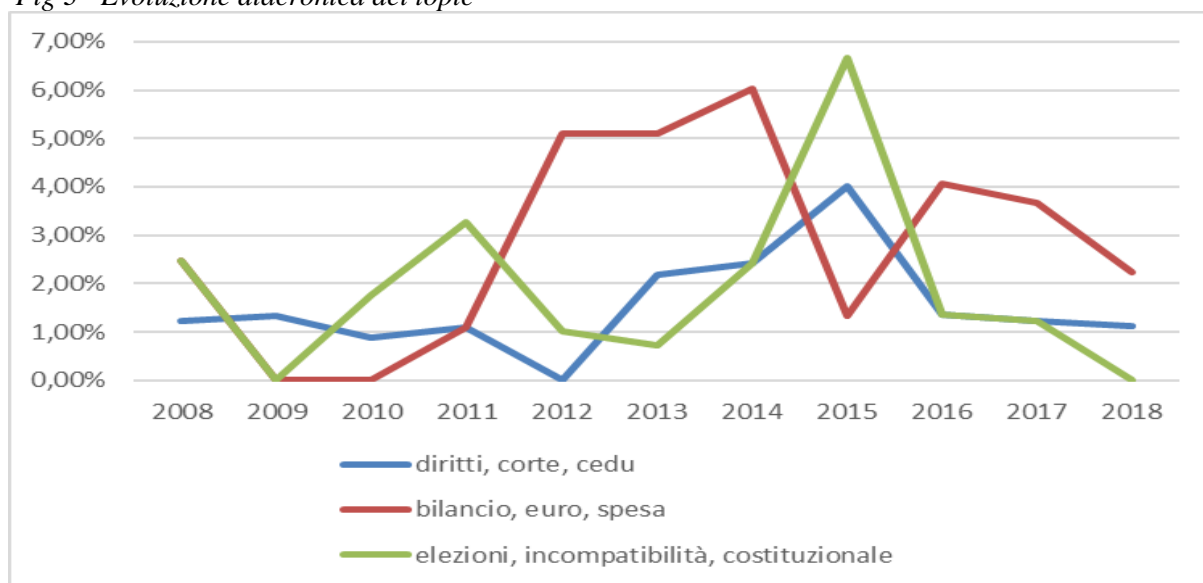
Fig. 2 Selezione del miglior modello



La figura 4 fotografa la frequenza dei topic estratti dal corpus di sentenze di illegittimità emanate dalla Corte costituzionale. In primo luogo, si può notare come la Corte abbia prestato particolare attenzione al topic lavoro. La tutela del lavoro, pubblico o privato, rappresenta un tratto genetico dell'attività della Corte in quanto rientra tra i principi fondamentali della Costituzione (Art. 1 Costituzione). In particolare, negli ultimi anni la Corte si è interessata al personale delle pubbliche amministrazioni ed ha orientato le proprie decisioni ad un sostanziale contenimento della spesa. A riguardo, è utile richiamare la sentenza 310/2013 in materia di congelamento degli stipendi dei professori universitari che è stato ritenuto necessario alla luce dell'obiettivo del contenimento della spesa pubblica. Si ricordi, inoltre, il recente intervento su una delle *key-policy* del governo Renzi come il Jobs Act di cui la Corte ne ha censurato il nucleo ovvero il contratto a tutele crescenti. In secondo luogo, sono molto frequenti i riferimenti al topic giustizia, che rappresenta un'area su cui la Corte si è espressa in maniera più decisa a partire dagli anni Novanta (Pederzoli, 2008). Di fatto, la cospicuità di questo topic segnala una forte influenza della Consulta sulle politiche penali, e dunque su un campo molto delicato in quanto va ad incidere sui principali strumenti che fissano le regole tra stato e cittadini sottoposti ad accusa penale. In terzo luogo, dall'analisi emerge un'elevata frequenza di topic legati alla caccia, area di legislazione che rientra in via esclusiva tra le competenze regionali. Si tratta di un dato molto interessante in quanto è da considerarsi sia come un riflesso dell'inefficiente riparto di competenze attuato dalla riforma del Titolo V della Costituzione che dell'assenza di una sede politica adatta a risolvere controversie che spesso si riconducono a frammenti di leggi regionali. Di conseguenza, sui tavoli dei giudici costituzionali giungono questioni che dovrebbero ritrovare la propria composizione in altre sedi causando l'inevitabile rallentamento dei lavori della Consulta. In quarto luogo, la tabella chiarisce ulteriormente l'impatto della riforma del Titolo V attraverso un richiamo diretto ai principi che regolano le relazioni tra Stato e regioni dopo la sua introduzione. In particolare, si evince il riferimento al principio di leale collaborazione che, disciplinato dall'art. 120 della Costituzione, ha rappresentato una delle principali innovazioni introdotte dalla riforma costituzionale. Esso mira a vincolare lo Stato e le regioni sul piano dell'amministrazione e rappresenta una delle "linee guida" nel disciplinare la relazione tra Stato ed Enti territoriali. In quinto luogo, possiamo apprezzare anche i riflessi e i limiti della riforma del Titolo V, che ha esaltato il ruolo arbitrale e di supplenza da parte della Corte durante la sua fase di implementazione. Di fatto, la riforma avrebbe dovuto condurre ad una maggiore autonomia delle regioni e ad un rimodellamento delle relazioni centro-periferia. Tuttavia, le difficoltà sono state numerose come testimoniano la crescita dei contenziosi tra Stato e regioni e le numerose sentenze di illegittimità della Consulta su leggi promulgate dai governi regionali. La scarsa razionalizzazione delle competenze tra Stato e regioni è legata in particolare agli

squilibri legati alle competenze “concorrenti” tra Stato e regioni previste dall’articolo 117 della Costituzione. In merito, si consideri la rilevanza dei topic relativi al settore dei trasporti e dell’energia. Inoltre, è molto interessante osservare come la nostra analisi suggerisca come il riordino degli enti pubblici quali province ed enti locali non sia passata inosservata agli occhi dei giudici della Consulta. Si pensi alla dichiarazione di illegittimità su una delle proposte più innovative avanzate dal Governo Monti che nel decreto-legge “Salva Italia” disponeva l’abolizione delle province. Infine, dall’analisi emergono le difficili relazioni con le regioni a Statuto speciale come la Sicilia e le province autonome di Trento e Bolzano su cui la Corte è intervenuta soprattutto in materia di adeguamento della spesa pubblica ai regimi delle regioni a Statuto ordinario durante gli ultimi anni.

Fig 3– Evoluzione diacronica dei topic



Guardando all’evoluzione dei topic in dimensione diacronica è utile soffermarci su alcuni topic che suggeriscono alcune delle tendenze più recenti della giustizia costituzionale e delle sue relazioni con l’arena politica. In primo luogo, possiamo notare una sostanziale stabilità dell’area bilancio e finanza pubblica lungo gli undici anni presi in esame. Essi sono da considerarsi una delle conseguenze più evidenti della crisi economica del 2008. Di fatto, la cospicuità dei topic bilancio e finanza mettono in luce la necessità dei diversi attori politici di prestare attenzione al debito pubblico, al rispetto del pareggio di bilancio e il divieto di indebitamento. Quest’ultimi sono stati inseriti nella Costituzione con la modifica dell’articolo 81, e dunque sono da considerarsi come aree su cui la Corte ha piena competenza e, di conseguenza, maggiori responsabilità (Rebessi e Zucchini 2018). Va sottolineato che la Corte ha storicamente prestato molta considerazione. Negli anni Ottanta, ad esempio, le sentenze 107 e 270 del 1988 chiarivano come la gestione dei tributi da parte dello Stato debba “essere assoggettato alla garanzia costituzionale della sua gestione”, e incitavano ad una più generale revisione del sistema tributario italiano. Negli ultimi anni, la principale differenza rispetto al passato è da individuarsi nell’enfasi posta dalla Consulta agli effetti sul deficit delle sentenze (Barsotti et al. 2016). L’attenzione all’equilibrio della finanza pubblica emerge, ad esempio, con la sentenza 10/2015 sulla cd. Robin Tax, su cui era stato stimato un buco di bilancio di sette miliardi di euro. La Corte ha quindi optato per una sentenza manipolativa in quanto una di tipo tout court avrebbe violato l’equilibrio di bilancio previsto dall’art. 81 della Costituzione e causato una nuova manovra finanziaria.

In secondo luogo, la Corte costituzionale ha posto particolare attenzione al tema dei diritti e delle libertà individuali, in modo da tutelare il rispetto dei diritti umani che rappresenta uno dei principi fondamentali della Costituzione. La loro tutela e sviluppo ha visto fin dalla sua introduzione un ruolo importante della Corte costituzionale, le cui decisioni hanno profondamente influenzato il decisore politico nei suoi orientamenti. Si pensi alla sentenza 80/2010 in materia di persona con disabilità, alla sentenza 138/2010 sui matrimoni omosessuali e 151/2009 sulla procreazione assistita. In particolare, la Corte ha sempre attribuito particolare importanza al diritto alla vita osservando che esso “assume una posizione privilegiata nell’ordinamento legale in quanto attiene alla vera essenza di quei valori supremi su cui si basa la Costituzione Italiana” (sentenza 1146/1988). La crescente tutela dei diritti civili e politici e poi di quelli sociali, è stata possibile grazie al ruolo dei giudici ordinari che sono divenuti sempre più protagonisti nell’attuazione del dettato costituzionale, e al supporto della giurisprudenza costituzionale europea. Di fatto, spostando la nostra attenzione dal livello nazionale a quello sovranazionale, possiamo notare che si vi sia un rapporto più stretto con la giurisprudenza della Corte costituzionale e le Corte Europea dei Diritti dell’Uomo (CEDU). L’elevata frequenza di questo topic rispecchia la crescente autorevolezza dei tribunali sovranazionali e il maggior dialogo tra giudici nazionali e sovranazionali negli ultimi anni.

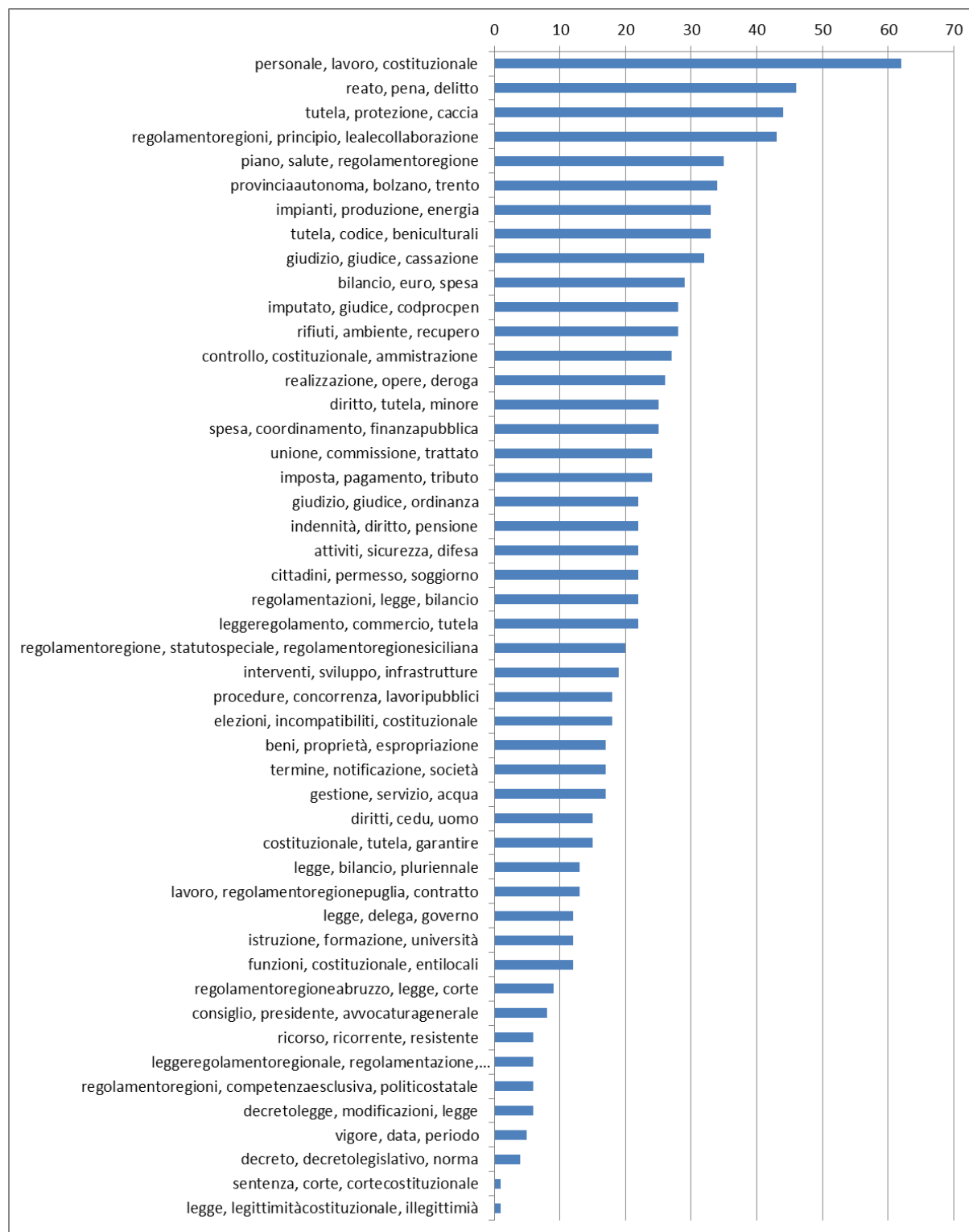
In terzo luogo, il topic relativo alla materia elettorale merita un approfondimento per le implicazioni che le decisioni della Corte hanno avuto sull’arena politica. Di fatto, la Corte costituzionale è sembrata ampliare le proprie capacità di regolazione dell’attività politica, anche attraverso il monitoraggio della condotta di candidati ed eletti, nonché delle regole per la loro elezione (Rullo 2020). La Corte ha mostrato una crescente attenzione alla verifica delle incompatibilità delle cariche e a bilanciare i diritti all’elettorato attivo e passivo. A livello regionale si ricordino la serie di sentenze che ha riguardato la regione Sicilia (143/2010, 277/2011, 294/2011) in materia di incompatibilità. A livello nazionale, si ricordino le sentenze 1/2014 e 35/2017 in materia di legge elettorale nazionale, che hanno sostanzialmente marcato lo scivolamento verso il terreno della discrezionalità, e profondamente impattato la forma di governo italiana. Con la sentenza 1/2014 la Corte sembrerebbe aver riportato in auge il sistema proporzionale ed introdotto il sistema delle preferenze per entrambe le Camere. Con la sentenza 35/2017, la Consulta ha sostanzialmente modificato la legge elettorale *Italicum*, argomentando la sua decisione sulla base della lesione del principio d’eguaglianza del voto, in quanto eccessivamente sbilanciata verso il maggioritario.

5. Conclusioni

Questo lavoro ha inteso cogliere l’opportunità di utilizzare tecniche di text mining allo studio di decisioni giudiziarie nell’abito delle relazioni tra corti e governo. In particolare, l’analisi delle sentenze di illegittimità ha fornito una panoramica sui temi cui la Corte costituzionale italiana si è interessata dal 2008 al 2018. Questa analisi quantitativa ha permesso di comprendere da nuove angolazioni lo studio delle relazioni tra giustizia e politica. La tecnica utilizzata ha permesso l’esplorazione del corpus in un’ottica di analisi del contenuto orientata a fini esplorativi, permettendo una categorizzazione dei testi e l’individuazione dei topic latenti all’interno di un corpus costituito da documenti molto lunghi e con un’elevata variabilità lessicale. L’analisi quindi ha restituito un quadro variegato che riflette innanzitutto una fase di profondo mutamento nelle relazioni tra giustizia e politica. Il lavoro ha inteso far luce sui topic su cui la Corte si è espressa dal 2008 al 2018. Sono stati approfonditi alcuni di questi temi che hanno fatto emergere particolari frizioni tra la Corte costituzionale e

Parlamento come il caso delle leggi elettorali. Allo stesso tempo, è stato messo in evidenza la diffusa attenzione al contenimento della spesa pubblica durante gli anni che hanno seguito la crisi economica del 2008.

Fig 4 – Topic Estratti



6. Bibliografia

- Amaturo, E., & Punziano, G. (2013). *Content Analysis: tra comunicazione e politica*. Ledizioni, Milano.
- Balakrishnan N., Nevzorov V. (2003). *A Primer on Statistical Distributions*, Wiley-Interscience.
- Barsotti, V., Carozza, P. G., Cartabia, M., & Simoncini, A. (2015). *Italian constitutional justice in global context*. Oxford University Press.
- Berry, M. W., Kogan, J (2010). *Text mining: applications and theory*, John Wiley & Sons.
- Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent dirichlet allocation, *Journal of machine Learning research*, 3(Jan), 993-1022.
- Blei, D. M., Lafferty, J. D (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 1(1), 17-35.
- Blei, D. M., Lafferty, J. D (2009). Topic models. In *Text Mining* (pp. 101-124), Chapman and Hall/CRC.
- Bolasco, S., De Mauro, T. (2013). *L'analisi automatica dei testi: fare ricerca con il text mining*, Carocci Editore.
- Buntine, W (2009). Estimating likelihoods for topic models. In: *Asian Conference on Machine Learning* (pp. 51-64). Springer, Berlin, Heidelberg.
- Chang, J., Blei, D. (2009). Relational topic models for document networks. In: *Artificial Intelligence and Statistics*, pp. 81-88.
- Charu R., Amit V., Bharath Y., Rama K., Kiran S. (2014). *API-FICATION*, Hcl Technologies, https://www.hcltech.com/sites/default/files/apis_for_dsi.pdf.
- Dyevre, A., Wijtvliet, W., & Lampach, N. (2019). The future of European legal scholarship: Empirical Jurisprudence. *Maastricht Journal of European and Comparative Law*, 26(3), 348-371.
- Faust, K. (1997) Centrality in affiliation networks. In: *Social networks*, 19(2), 157-191.
- Griffiths, T. L., Steyvers, M. (2002). A probabilistic approach to semantic representation. In: *Proceedings of the annual meeting of the cognitive science society*, Vol. 24, No. 24.
- Griffiths, T. L., Steyvers, M. (2003). Prediction and semantic association. In: *Advances in neural information processing systems*, pp. 11-18.
- Griffiths, T. L., Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl 1), 5228-5235.
- Guarnieri C., Pederzoli P. (2017), *Il sistema giudiziario*, Il Mulino, Bologna
- Hallach, H. M., Murray, I., Salakhutdinov, R., Mimno, D. (2009). Evaluation methods for topic models. In: *Proceedings of the 26th annual international conference on machine learning* (pp. 1105-1112). ACM.
- Heinrich, G. (2005). *Parameter estimation for text analysis*. Technical report.
- Hirschl R., (2008) "The judicialization of politics", In *The Oxford Handbook of Law and Politics*, ed. Keith E. Whittington, R. Daniel Kelemen, and Gregory A. Caldeira, Oxford , UK: Oxford University Press, pp. 119-141
- Hofmann, T. (1999). Probabilistic latent semantic analysis. In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pp. 289-296, Morgan Kaufmann Publishers Inc.

- Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42(1-2), 177-196.
- Lenci, A., Montemagni, S., Pirrelli, V., & Venturi, G. (2009). Ontology learning from Italian legal texts. *Law, Ontologies and the Semantic Web*, 188, 75-94.
- Lelli, S. M., Sigismondi, I., Turra, R., Pedrazzi, G., & Rago, S. (2011). Estrazione automatica di informazioni dai testi delle sentenze. In Fabri M. (ed.), *L'altra e-justice*, Clueb, Bologna, pp. 43-63.
- Losito G., 1993, *L'analisi del contenuto nella ricerca sociale*, Milano, Angeli.
- Montemagni, S., Peters, W., & Tiscornia, D. (2010). *Semantic Processing of Legal Texts*. Springer-Verlag Berlin Heidelberg.
- Musella F. (2018), *Political leaders beyond party politics*, Palgrave/Macmillan, Cham.
- Neapolitan R. E. (2003). *Learning Bayesian Networks*, Prentice-Hall, 58 (4), pp. 1064-1082.
- Nevzorov, V. B., Balakrishnan, N., Ahsanullah, M. (2003). Simple characterizations of Student's t_2 - distribution. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3), 395-400.
- Pederzoli, P. (2008) *La Corte Costituzionale*, Il Mulino, Bologna.
- Steyvers, M., Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7), 424-440.
- Steyvers M., Griffiths T. (2007). Probabilistic topic models. In: *Latent Semantic Analysis: A Road to Meaning*, eds. T. Landauer, D. Mcnamara, S. Dennis, and W. Kintsch, Lawrence Erlbaum, page 427.
- Puschmann, C., & Ausserhofer, J. (2017). *Social Data APIs: Origin, Types, Issues*.
- Rebessi, E., & Zucchini, F. (2018). The role of the Italian Constitutional Court in the policy agenda: persistence and change between the First and Second Republic. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 48(3), 289-305.
- Rullo, L. (2020) *L'espansione della Corte costituzionale nell'arena politica (2008-2018)*, Tesi di dottorato, Dipartimento di Scienze Sociali, Università di Napoli Federico II.
- Sagri, M. T., Agnoloni, T., & Bacci, L. (2014). Tecniche di estrazione terminologica e classificazione automatica di corpora giurisprudenziali. *Informatica e diritto*, 23(1), 41-64.
- Sigismondi, I., Turra, R., Pedrazzi, G., & Rago, S. (2011). Text mining sulle sentenze: nuove possibilità per un sistema di monitoraggio della giustizia. In Fabri M. (ed.), *L'altra e-justice*, Clueb, Bologna, pp. 25-42.
- Shi, L., Zhong, H., Xie, T., & Li, M. (2011, March). An empirical study on evolution of API documentation. In *International Conference on Fundamental Approaches To Software Engineering* (pp. 416-431). Springer, Berlin, Heidelberg. Author A. (2000).
- Tipaldo, G. (2014). *L'analisi del contenuto ei mass media*. Bologna, IT: Il mulino.
- Vallinder T., Tate C.N., eds. (1995), *The global expansion of judicial power*, New York University Press, New York.