

Dynamic Analysis of Archives Contents: the Case of the ISNIE/SIOE Intellectual History

Eric Brousseau*, Bruno Chaves**, Svitlana Galeshchuk**

University Paris-Dauphine, PSL Research University

*Governance & Regulation Chair and Governance Analytics

**Governance Analytics

Very First and Incomplete Draft

Abstract

The purpose of this paper is to present the methodology we have been developing to explore the thematic structure of a collection of texts and its dynamic without an appropriate a priori understanding of the relevant categories and concepts. Since we believe that this methodology is particularly promising for studying the organizational and institutional phenomena thanks to relevant archives or digital tracks, we chose to focus our attention on the “archive” of the Society for Institutional and Organizational Economics (SIOE) that celebrated its 20th anniversary in 2018. This archive comprises the abstracts of more than 3 000 papers that were presented at the 21 annual conferences of the society organized between 1998 and 2018. The dynamic topic modeling of this data highlights two major trends. First, there is over time a reversal of fortune between the study of institutions and the analysis of organizational arrangement that have been becoming dominant from the 2010’s. Second, within each main domain of investigation, a shift from a broad approach contrasting alternative orders (e.g. market vs. hierarchy, private vs. collective property), to a micro-analytical approach of mechanisms (either contractual, organizational or institutional) can be observed.

1. Introduction

For the last decade, academics and practitioners have been getting access to an increasing volume of data as many aspects of individual life and collective activities have become digitized, and also because the costs of digitizing paper archives spurs interest in exploiting them to manage evidence-based policies and strategies. This led to forging the term “Big Data” and broadening a set of methods to analyze large volume of rich, complex data. Such developments do not only transform the process of decision-making (due to automation based on machine/algorithmic learning) but also generate new knowledge by virtue of the ability to explore rich data set and to detect underlying structure that might refer both to unknown phenomena and causal relationships. Having said that, frequently “big data” refers not to formatted/quantitative information but to files containing images, video, and text; i.e. information in an unstructured form. Treatment of these types of data aims in particular at shaping emerging concepts and causal relationships, and also at discovering new phenomena. Indeed, for the well-established categories and analytical models, all kinds of structured/codified data are available to test causality relations. The exploitation of data in its natural form — e.g. natural language — clearly corresponds to the purpose of exploratory field studies and interviews that are performed in natural and social sciences to explore the shape of a phenomenon before attempting to implement more systematic tracking and measuring. The ability to explore a large set of data without prior

understanding is a potential source of additional discoveries as it helps reduce the individual cognitive biases of the observer and lessens the role of specific conditions of the exploratory phase in field studies.

Methods of systematic exploration represent a significant opportunity for the research community in social sciences, and in particular for those interested by institutions and organizations (see Prüfer and Prüfer, 2019), since they potentially allow to identify underlying structures that are beyond the immediate cognitive capabilities of researchers. To draw a parallel with network analysis, the development of systematic methodologies to simply identify networks structures in large data sets dramatically changed our capability to describe a wide set of social phenomena, and to explain them. Even the ability to observe and describe networks allowed to shift the understanding of specific phenomena (like the role of key intermediary in systems of exchange) from one set of explaining variables (e.g. their endowment in the context of their local community) to network phenomena (e.g. they are pivotal/key nodes in the interconnexion between heterogeneous networks). The systematic analysis of large archives (initially in digital form or not) represents exactly the same potential for research in institutions and organizations. Indeed, the characteristic of many of the later is to record their activity. Before the big data fashion, social sciences heavily relied upon these structured archives to study formal organizations. Accounting systems and tracks of financial operations have been extensively used for that purpose. However, the unstructured archives, like the written reports, the bureaucratic memos, the instructions exchanged within hierarchies, the minutes of meetings, the correspondence remain either systematically unprocessed or their exploitation targets some specific topics, for a small set of organizations, and are performed manually by researchers, especially in history, sociology or political science. Today such unstructured archives are digitized and can be analyzed more systematically. In particular, their content can be scrutinized with text mining methods to understand how it is structured and how it evolves over time. This is of course only a first step to use the analysis of the content of an archive either to characterize an organizational phenomenon (like the evolution of the behavior of an organization) or to explain it (like the evolution of the agenda/policy of a group).

The purpose of this paper is to present the methodology we have been developing in the context of the Governance Analytics initiative¹ to explore the thematic structure of a collection of texts and its dynamic without an appropriate a priori understanding of the relevant categories and concepts. Since we believe that this methodology is particularly promising for studying the organizational and institutional phenomena thanks to relevant archives, we chose to focus our attention on the “archive” of the Society for Institutional and Organizational Economics (SIOE) that celebrated its 20th anniversary in 2018. This archive comprises the abstracts of more than 3 000 papers that were presented at the 21 annual conferences of the society organized between 1998 and 2018

Originally established as the International Society for New-Institutional Economics (ISNIE), the aim of SIOE is to develop a better understanding of governance phenomena through innovative and interdisciplinary, while anchored in economics, research. The underlying assumption is that most economic performances are driven by organizational and institutional factors, and that the political-economy of governance systems and their dynamics has to be understood to trigger organizational changes or to manage policies. Moreover, since the understanding of these phenomena requests knowledge initially anchored in other disciplines ranging from anthropology to management, the society promotes trans-disciplinarity. While the field benefitted from some recognition in the early

¹ Governance Analytics is a research group that has brought together researchers in Computer Science, Economics, Management, Political Science and Sociology who work on the analysis of governance issues. It aims to build capabilities and methods to acquire and process data for collective decision-making mechanisms, institutional dynamics, and the functioning of markets and organizations. <https://www.governanceanalytics.org/en>

1990's, in particular through the awarding of the Nobel prize in economics to Ronald Coase (1991) and Douglas North (1994) — the two founders of ISNIE —, it remained a niche in economics. In addition, the transdisciplinary nature of IOE scattered its visibility through a great deal of epistemic communities: law and economics, economic history, development, industrial organization, finance (corporate governance), etc. Thus, the establishment of the society was clearly aimed at better structuring the field and the underlying research community, in particular by organizing an annual conference, in order to promote its approach and the study of its favorite topics in economics. The society therefore made efforts to attract scholars both from all fields in economics and from other disciplines in its annual conference.

Interestingly, the society decided not to launch its own journal in order to incite its members to publish in the existing journals in the aim of reaching a large audience, both within the economic profession and in other disciplines in social sciences. Thus, the main achievement of the SIOE since 1997 has been the recurrent organization of an annual conference gathering a collection of selected papers of contributors to the field. We can expect that, on the one hand, the researchers that want to be recognized contributors to the field are likely to propose their recent papers to the SIOE audience both to get feedbacks and to promote their work. On the other hand, the selection committee is incited to select the papers that it considers as relevant for the field and that meet a minimal threshold of quality. While it is obviously a biased sample of the research performed on organizations and institutions in economics — because some researchers on these topics might consider themselves as involved in another field like industrial organization or development, we can consider that the study of the set of papers given at SIOE provides a vision of what is studied in the field on how it evolved over time. Indeed, even if some researchers never attended the SIOE meeting but contributed significantly to the development of knowledge in the analysis of institutions and organizations, the topic they have been studying should have influenced the work of those who presented their work at SIOE. By investigating the evolution of the SIOE conference focus and themes for the last 20 years using dynamic topic modelling and other text mining and machine learning methods, our paper provides both a contribution to the history of the SIOE in particular, and to the analysis of the structure of the institutional and organizational economics and its evolution over time.

Such an intellectual history is however only the ancillary goal of this paper, which main contribution is methodological by highlighting on the basis of the exploitation of a specific corpus how data mining approaches may be exploited to conduct a profound analysis of topics and main concepts in a textual archive.

The rest of the paper is organized as follows: Section 2 describes the data and its pre-processing. Section 3.1. provides details of our methodology for studying the evolution of topics as evidenced from the SIOE conference abstracts over 20 years using dynamic topic modelling. It also explains how our approach may be evaluated using model coherence methods. Section 3.2 amplifies on semantic networks. Section 3.3 elaborates on text summarization. Section 4 presents our main results and interpretation, and section 5 concludes with our findings and suggestions for future works relying on this methodology.

2. Data

Documents in our study correspond to the abstracts² from the SIOE conferences over the last 21 years.

² In this paper, the word “abstract” is used interchangeably with “document”

2.1. Collecting the data

We collected the titles and abstracts of all ISNIE/SIOE conferences from 1998 (the year of the foundation of the society)³ to June 2018 (21 years). For the period of 2008 to 2018 this data has been extracted from the MySQL database of the society's website to a CSV (comma-separated values) text file⁴. However, prior to 2008, the conference program was edited manually by the conference secretariat using MS Word files. Consequently, for this period, we had to copy and paste the titles and abstracts manually from the MS Word files to the CSV file. A total of 3228 titles and abstracts have been collected but only 3225 were effectively used because 3 abstracts had only few words (less than 30).

Table 1 : Number of titles and abstracts by year (collected / used)

1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
50/50	87/87	90/90	106/106	93/93	71/71	176/176	157/157	175/175	176/176	111/111
2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
217/217	114/114	175/174	152/152	238/238	158/158	195/195	255/254	241/240	191/191	3228/3225

2.2. Preprocessing the data

The data has been preprocessed with the Spacy and Gensim Python libraries⁵ following the steps listed below:

- Normalization to US English by replacing UK English words by their US English equivalents⁶.
- Removal of usual English “stop words” (*these are mainly common words with no semantic interest*).
- Removal of corpora-specific “stop words” (*e.g. article, abstract, paper, etc.*)
- Removal of useless parts of speech for topic modeling (*articles, adverbs, etc.*).
- Word lemmatization (i.e. grouping together the inflected forms of a word so they can be analyzed as a single item)⁷
- Identification of the corpora-frequent bigrams such as “transaction costs”, “property rights”, “human capital”, etc.

3. Methodology

Comprehensive data analysis requires integration of complementary text mining methods. Hence, the methodological outline to address our research question must imply the combination of extant cutting-edge approaches to investigate the thematic structure of available texts in depth. Our research pipeline defines its three principal steps as follows:

³ Actually, an inaugural conference was organized at Washington University in St Louis in September 1997 to launch the Society that was established a few months later. The second conference was organized in Paris in September 1998. Only (invited) plenary talks were given at the inaugural conference.

⁴ These abstracts have been uploaded by the presenting authors inside the society extranet.

⁵ <https://spacy.io/> and <https://radimrehurek.com/gensim/>

⁶ This is particularly important in the context of international conferences.

⁷ For example, in English, the verb 'to walk' may appear as 'walk', 'walked', 'walks', 'walking'. The base form “walk” is the lemma for the word. See: <https://en.wikipedia.org/wiki/Lemmatization>

Step 1: Topic modelling for the collection of abstracts. This step returns n number of topics and a list of words representing each topic. It also groups the abstracts assigning the topics to each document based on the topic's distribution.

Step 2: Semantic Networks for each cluster of abstracts resulting from Step 1 to investigate their principal concepts.

Step 3: Text summarization for each cluster of abstracts resulting from Step 1 to get a better understanding of their core ideas.

3.1. Topic modelling

3.1.1. *Static topic modelling*

Recall that we use the corpus of the SIOE abstracts over 20 years to infer the conference thematic structure and its evolution. Among the available methods, probabilistic topic modelling demonstrates, perhaps, the most promising capacity to address this challenge. Topic modelling includes a group of related machine learning methods useful for natural language understanding. It develops a probabilistic model for a collection of documents (abstracts); the model provides insights into the topics discovered in the text and helps uncover inter-relationships between these topics by identifying commonly used terms.

Over the last decade, topic modelling transformed into one of the main technics in text analysis and summarization. This method comes from the family of unsupervised approaches used for obtaining hidden thematic structures in textual material. It has been employed by researchers in the number of domains (see Kaplan and Vakili 2012; Vulić et al. 2015). Jaccobi, van Atteveldt, Welbers (2016) use topic modelling in journalism, in particular for analyzing trends and patterns over news content. Hu et al. (2014) apply this technic for the corpus of patent documents. Despite its generally good performance over aggregated collection of documents, the approach in its vanilla version (static topic modelling) cannot be employed to monitor the topic evolution over time. This drawback may be tackled by using the dynamic topic modelling.

3.1.2. *Dynamic topic modelling*

Salient topics in discipline and common terms in a topic typically change over time. In their paper, Blei and Lafferty (2006) propose employing continuous distribution in order to follow this evolution. The researchers focus on monitoring the topic transformation with the dynamic topic modelling algorithm. We adopt the same approach to study the dynamics of thematic change at the SIOE conferences. However, we also try to address the shortcomings arising from unsupervised learning (i.e., the choice for a number of clusters, topics) by evaluating the topics' coherence with available quantitative methods.

Dynamic topic modelling captures evolution over time by partitioning documents into time slices; the distributions of topics and the distribution of words in a topic are considered to evolve over time. The posterior distributions are obtained from the likelihood function and the priors at the previous time slice. Each document in a dynamic topic model is viewed as a mixture of unobserved topics. Inference involves learning the other parameters of a generative model based on words observed in documents over a sequence of time slices. Based on distributional assumptions, the priors and the posteriors do not belong to a conjugate family of distributions for which the parameters can be conveniently updated. Since exact estimation of the parameters is intractable, we employ Variational Kalman Filtering for the approximate inference of these parameters.

Using the notation of Blei and Lafferty (2006), the evolution of topics over the time may be formally described as follows:

Let:

α_t be the topic distribution in documents at time t , and

$\beta_{t,k}$ be the distribution of words for topic k at time t .

$Z_{t,d,n}$ represent the topic for word $W_{t,d,n}$ (indicates the n th word in document d at time t).

Under dynamic topic model the evolution of these parameters may be specified by their conditional distributions:

$$\alpha_t | \alpha_{t-1} \sim N(\alpha_{t-1}, \sigma^2 I)$$

$$\beta_{t,k} | \beta_{t-1,k} \sim N(\beta_{t-1,k}, \delta^2 I)$$

The generative process at time t may be specified as:

For all topics k , draw topics $\beta_{t,k} | \beta_{t-1,k} \sim N(\beta_{t-1,k}, \delta^2 I)$

Draw $\alpha_t | \alpha_{t-1} \sim N(\alpha_{t-1}, \sigma^2 I)$

For each document d :

Draw $\eta_{t,d} \sim N(\alpha_{t,k}, a^2 I)$

For each word:

Draw topic $Z_{t,d,n} \sim \text{Mult}(\pi(\eta_{t,d}))$

Draw $W_{t,d,n} \sim \text{Mult}(\pi(\beta_{t,Z_{t,d,n}}))$

Here $\pi(x)$ maps the natural parameterization x to the mean parameterization as:

$$\pi(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

The only observable is $W_{t,d,n}$. The inference involves learning the other parameters of the model from these observations.

A Python wrapper for Dynamic Topic Models is available in the package *gensim* (<https://radimrehurek.com/gensim/models/dtmmodel.html>). We use this package to develop our code script for topic modelling.

3.1.3 Number of topics

Topic modelling has its disadvantages. One of such drawbacks is an unknown number of “true” topics in text. We use the coherence score to tackle this problem and estimate the optimal number of topics. Following Roeder, Both and Hinneburg (2015) we use the 4-stage evaluation when the coherence score is computed as the mean pairwise Cosine similarity between vectors corresponding to the most common terms. These vectors are inferred via the Word2Vec neural network model⁸. Fig. 1 plots the coherence score for the number of topics in a range [2:14]. The visualization shows the diminishing returns if more than 8 topics are deduced.

⁸ Word2Vec model is explained in details in Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

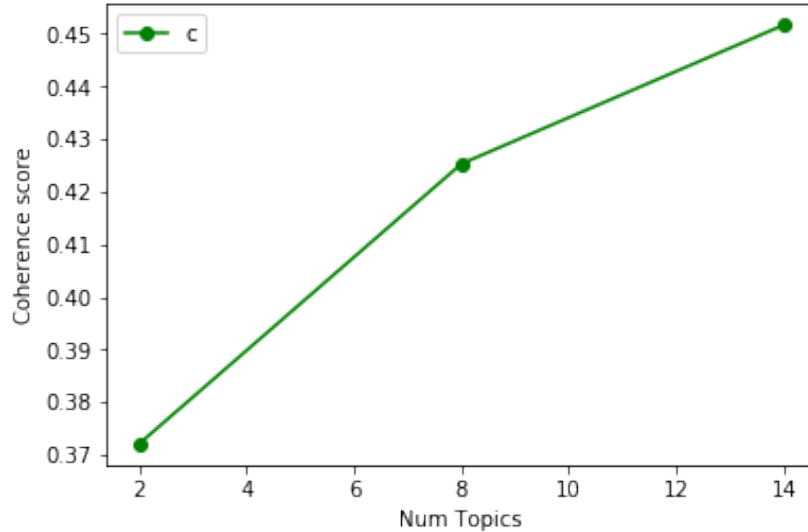


Fig. 1. Optimal number of topics

3.2 Semantic Networks

Concepts, theories may be defined through the number of keywords that represent their meaning and functions. The more aspects a concept embraces, the longer is a list of its key terms. For example, the word “fruit” encompasses the idea of “tree”, “vitamins”, “food”, “colorfulness”, etc. Through these terms, “fruit” is indirectly connected with a concept of “healthy life” and the like. By putting these and other meaningful connections together we may create a network where the nodes are the words and the edges denote the proximities between the nodes. Such graphs displaying the contextual proximity between lexical units define semantic networks. The nodes with the largest number of connections stay in the center of the graph.

Each node (word, entity) is associated with weighted edges linking it to the other terms that co-occur within a determined sliding window of words in a text. For example, if this window equals 2 then only terms that co-occur adjacently will make an edge. In our methodological set-up, the window is set to 10 (see Appendix 2)

3.3 Text summary

Topic modelling returns a list of words with respect to the topic. However, this set of terms does not necessarily provide a clear understanding of the gist for the documents united by the same topic. What if we could generate some pithy summary of corpora retaining their core meaning? Fortunately, mathematical and statistical foundations allow to formalize and implement automated document summarization approaches.

We have tried to implement several statistical methods including the Latent Semantic Analysis.⁹ However, the TextRank method provides the most comprehensive results. It builds on the so-called PageRank approach, which used to be employed as a core algorithm of the Google search engine: web pages contain the links to the other web sites interconnecting them; in graph model the web pages may represent the nodes with the edges corresponding to the links among them. Moreover, when one web page connects to another one in the graph, the former basically casts a vote. The size of nodes

⁹ The core idea behind the Latent Semantic Analysis is that it exists some latent structure among the words related contextually. These words must also correlate in the same singular space.

depends on the web page importance. This measure is computed not only on the basis of the number of votes but also on the importance of voting web pages that are connected to it.

Concerning the TextRank algorithm, when summarizing corpora, we have sentences or phrases as the nodes of our graph. The nodes may share multiple connections, however, unlike the PageRank algorithm, these connections will depend on weight coefficient such that the weight indicates the strength of the link between the vertices. The term frequency–inverse document frequency (TF-IDF)¹⁰ provides a measure for the weight coefficients. When the graph is constructed, the algorithm chooses the top k sentences as an output summary.

The output of our treatment is displayed in Appendix 3

4.RESULTS

4.1. Graphic Representations of SIOE Topics and their Evolutions over Time

Figures in Appendix 1 depict the results of dynamic topic modelling applied to the collection of the SIOE abstracts. Each graph plots the top words dynamically characterizing a topic over time. The list of terms helps assign a title to a corresponding topic.

Recall from the generic dynamic topic modelling process (section 3.1.2) that the method returns not only the key terms per topic but also the thematic distribution for each document. This quantitative measure allows to compute the average score of topical presence per year and to observe its evolution.

We use two approaches to estimate the topic distribution over the collection of abstracts per year: the mean method and the ratio method. The former sums the mean of each topic per document and divides it by the number of documents (Figure 2). The latter first filters only the topic with the maximal distribution (maximal presence) for each document grouping the abstracts by subject. Then it computes the average of each topic per year (Figure 3). Both figures allow to visualize the evolution of topics estimated by the two approaches described. The lines are smoothed with 2-order moving average to capture better the trend for each topic.

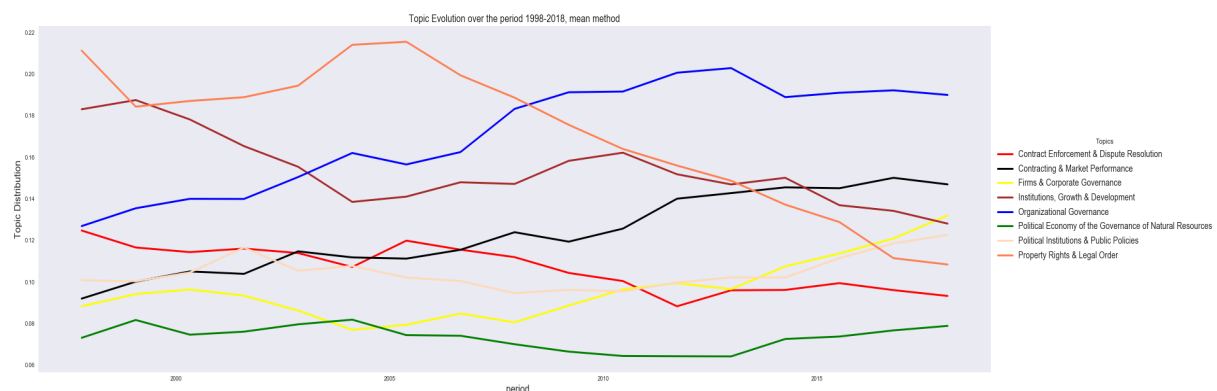


Fig. 2. Distribution of the IOE topics over time / means method

¹⁰ TF-IDF equals to the product of term frequency (the frequency of the word in a document) and inverse document frequency (the total number of documents in our corpus divided by the document frequency for each term and then applying logarithmic scaling on the result)

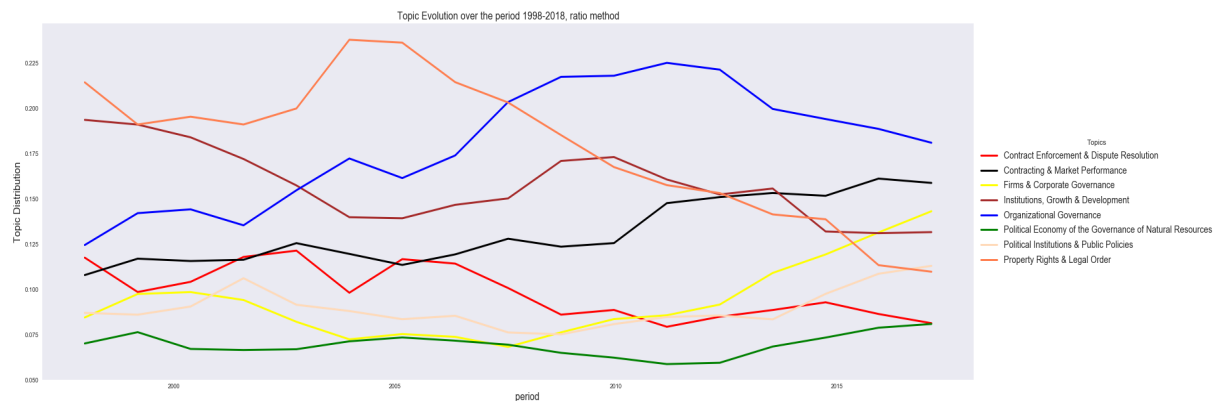


Fig. 3. Distribution of the IOE topics over time / ratio method

Appendix 2 includes the semantic networks constructed over 8 clusters of abstracts which contribute the most to the corresponding topic. These groupings are based on the maximal distribution (maximal presence) of a particular topic in an abstract. The key topical words keep the highest centrality in the network.

Appendix 3 lists the most important sentences/phrases over the cluster of abstracts which contribute the most to the corresponding topic.

4.2. Topic Analysis

To interpret the various topics identified by the dynamic topic modelling, we rely upon the combination of results of three methodologies illustrated by the figures in Appendix 1, 2 and 3. Note that the title for each topic is given by the analyst on the basis of such interpretation, while the identification of the most frequent significant words (appendix 1), of the clusters of contributing words to each category (appendix 2), and the summaries (appendix 3) are machine generated. In practice, the analysis of the most frequent key-words (appendix 1), allows the analyst to identify the central topic of each cluster (as contracting, development and institutions, etc.) but that the analysis of the web of clusters of words (appendix 2) allows to refine the analysis because the most frequent keywords as contract are present in several topics. The automated summaries (appendix 3) are useful to confirm the categorization deduced from the first two efforts of interpretation. It is however less useful than the two first categories since the generated summary might result into a text that is not fully consistent because it captures key sentences that are characterizing documents with similar focus, while not being able to check the internal consistency of the resulting “abstract”. It is however a useful check.

It is also useful to remind the reader that documents (SIOE abstracts) might pertain to several categories; i.e. deal with more than one topic. Moreover, the methodology allows to give a “weigh” to each topic for each document. Hence the methodology really allows to analyze the ‘content’ of each document and does not depend upon self-classification by the author (who select keywords and tracks along which his paper will be assessed and then presented) or by the decision of the conference organize to group papers in topical sessions.

As pointed above, the analysis of the 3225 of papers given at SIOE conferences for 21 years highlight the persistence of 8 main topics dealt with by the researchers in the field. They are presented below in the order automatically generated by the dynamic topic modelling (which however is meaningless):

“Contracting and Market Performance” refers to the papers dealing with the process of contracting, the performance of the resulting contracts and the consequence of these contractual relationships for the performance of markets; especially the labor market, markets dealing with specific risks (insurance and healthcare) and intermediary markets among firms (buyer-supplier relationship).

“Institutions, Growth and Development” refers to the papers dealing at the macro level of the relationship between formal and informal institutions (in the spirit of North) and economic performance and in particular growth through various vectors; in particular the level of trust and the ability of societies to accumulate human capital and foster investment. The question of transition pertains to this topic as well as the vector of trust; in particular the inclusiveness of the socio-economic regime and the credibility of the legal and political institutions.

“Political Institutions and Public Policies” refers to the papers in the spirit of the political economy dealing with the actual behaviors of public decision makers and their ability to influence citizens and economic agents behaviors given the political regime (democracy vs. authoritarianism) and the related legitimacy and fairness of public decision makers, and given the incentives and biases in the way the government and the public bureaucracy perform (e.g. corruption, lobbying, political cycle, etc.).

“Property Rights and Legal Order” refers to the research for which the law and the property right regimes are central. The legal order is clearly envisaged here in the perspective of the incentives of the residual claimant of rights and their ability to coordinate/contract among each other. The de facto regime of rights is affected both by customs/inherited social arrangements (e.g. commons, farming regime), and by public policies and regulations.

“Contract enforcement and dispute resolution” refers to the contributions focusing on the credibility, fairness, foreseeability of the contractual order given the nature and performance of the mechanisms taking in charge dispute resolution. The focus is clearly on legal enforcement, the performance of courts and the behavior of judges. However, arms-length relationships and reputation effects are also in the scope.

“Firms and Corporate Governance” refers to the papers focusing on the firm as an economic agent and focusing on the main factors that explain their behaviors and steer their performances. The focus is on their financial health and financial independence. The way the board of director is structured and influenced is highlighted; and two categories of agents of influence appear to be central the policy makers and the banks.

“Organizational Governance” refers to the papers analyzing alternative to market governance in the tradition of Williamson, while in the same time departing from it. Indeed, other dimensions than the hierarchical dimension of alternatives to markets are explored. The notion of cooperation and innovation are central. The focus is both on contractual and organizational mechanisms. There is a clear interest for the collective governance regimes (professions, industries) and for cooperation and alliances. Also, the figure of the entrepreneur is present in this cluster.

“Political economy of the governance of natural resources” refers to the papers taking a ‘fundamental’ approach of the interplay between the institutional regimes (a combination of the property right regime, the degree of discretion and benevolence of governmental intervention, and the rule of law) and the efficiency in exploiting natural resources (with a focus on water and land). Since the retroaction loop between the two is considered, the historic institutional heritage drawn from natural conditions/socio-political shocks like colonization are under scrutiny

4.3. Dynamic Analysis

There are two ways of analyzing the evolution of the research agenda over time. First, for each topic/domain, the graphs of Appendix 1 highlight sometime significant evolution of the frequency/weight of some top keywords, that might be interpreted in an evolution of the way to consider the related domain over time. Second, Figure 2 and 3 (above) highlight contrasted evolution of the relative frequency of topic over time, pointing out an evolution in the focus of the research community.

The interest for the topic **“Contracting & Market Performance”** has been steadily growing over time. Its frequency rose from around 10% in the early 2000’s to nearly 15% in 2018, where it became the second most studied topic in the conference as compared to a 5th position in the 2002 ranking. That said the growth of the interest for this topic has slowed since 2013. The analysis of the evolution of the most the significant words (Appendix 1) highlights an evolution of the approach that became more focused on contractual design rather than on related market performance, and also more focused on contracting parties’ behaviors.

“Institutions, Growth and Development” was the second most popular topic in the early 2000s, but its relative importance felt dramatically until 2007 (roughly from 19% to 14%) where it then reached an apparent threshold oscillating between 14 and 16% until 2015. It felt against since then to a possible new threshold of 12-13%, which resulted into losing its rank into the top 3 topics. Also, the approach evolved, and became probably more focused and micro founded as suggested by the falling influence of ‘country’, ‘institution’ and ‘trust’ in the determination of this cluster.

“Political Institutions & Public Policies” has been experiencing overall a “steady state” oscillating around 10% over time with highs around 2005 and 2018 and lows in the early 2000s and circa 2011-2012. It seems that the approach of the topic evolved, with a constant drop of the focus on political institutions and an increasing interest for the design of public policies, that seems to drive the more raise of the importance of this topic from 2015.

“Property rights and legal order” was the most influential topic from the creation of the society to the early 2010s when it started to be surpassed by **“Organizational Governance”**. In fact, its frequency started to drop in 2008 when it reached a peak of nearly 22%. Between 2008 and 2018 its “market share” was divided by two and the fall has been steady. While it remains a significant dimension of the interest of the researchers involved in the conference, it ranks now only 6th among the 8 topics. Also, the reference to property and property rights dramatically went down, suggesting that the interest is now more on the vectors of the distribution of decision rights, risks, incentives.

“Contract enforcement and dispute resolution” is a topic characterized by a progressive, while limited, erosion of its influence over the period. It ranked fourth and weighed around 12% in the early 2000s and lost a third of its influence overtime, becoming the second last popular topic. It seems that the focus also evolved, being less centered on contract enforcement per se and more on dispute resolution and judicial ordering.

“Firm and corporate governance” remained one of the less studied topics until 2010, when it started to become central in the research agenda of conference participants almost doubling its frequency over time (from below 8% in 2010). In 2018 it ranked third among the 8 topics. The evolution from 2010 is also characterized by the increasing centrality of the word firm, highlighting a growing interest for the drivers of the behaviors of corporations.

“Organizational Governance” has experienced the most significant growth over the period starting from a third rank and a 12,5% level to the first rank and with a little less than a 20% level. It seems to have reach a maximum circa 2013-2014 and has been steady from that peak at a marginally lower level. This is however by far the most dominant topic dealt with in SIOE conference papers. Interestingly, also the growth of the topic influence came with a deepened centrality of the word governance and a decreasing influence of the notion of transaction, suggesting that the focus switch from a transaction cost economics approach of governance to a broader understanding of alternative governance regimes.

The **“political economy of the governance of natural resources”** topic has experienced a steady state in terms of weigh in the conference over the year. It has always ranked last, and it is present at a level of around 7% of the corpus. It however persisted over time. That said, it seems that the approach of the topic switched from a focus on the specific nature and centrality of natural resources to the importance of the governmental role in the managing them.

Overall the dynamic topic modeling of the SIOE abstract over time allows to highlight a set of trends that are not totally surprising for the scholars in the discipline, but that present the advantages of being measurable and comparable. Not only the evolution of the weight of topics over time characterize the general trend of the all corpus, but the analysis of the drivers of the topics building allows to identify an evolution of the approach of each of them. In short, it seems that two major trends are highlighted. First, there is clearly over time a reversal of fortune between the study of institutions and the analysis of organizational arrangement that have been becoming dominant in the last period (Organizational Governance, Contracting and market performance) and also are characterized by a raising trend all along the 2010s (Firm & corporate governance). Second, within each topic, it seems that there is a strong trend to switch from a broad approach contrasting alternative orders (e.g. market vs. hierarchy, private vs. collective property), to a micro-analytical approach of mechanisms (either contractual, organizational or institutional).

5. CONCLUSIONS

Two conclusions can be drawn from our analysis. The first is methodological, the second refers to the intellectual history of IOE.

From a methodological viewpoint, the fact that our analysis of trends in the focus and the methodological preferences of the researchers in IOE are not surprising for those who know the field. It is important to highlight that this result is derived from a systematic algorithm-driven analysis. A lay-analyst of the corpus would have come to the same general conclusions, even if the ‘translation’ in a sound analysis of the history of this research sub-field and community requests some understanding of the topic they are dealing with. For us, this result highlights the power of the methodology to scientifically and systematically analyze the content of a corpus of text in dynamic. Moreover, beyond the general feeling that any scholar in the field might have detected some of the topics, the detailed and significant (dynamic) mapping we have presented is clearly beyond the reach of a traditional human based text analysis.

Second, when dealing with the specificity of the history of New-Institutional Economics and Institutional and Organizational Economics, it is interesting to note that the evolution pointed out above reflects clearly an appropriation by the economic profession, broadly speaking, of the issues, questions and main concepts developed and brought forward by the founding fathers of the society, namely the Nobel laureates (Coase, North, Williamson and Ostrom) in the field, but also some influential scholars like Aoki, Barzel, Demsetz, or Omstron. As expected by Coase in a lecture at the AEA in 1991, these topics were progressively appropriated by a new generation of scholars, which in addition relied upon both the development of the theory (game and incentives theories in particular) and of the applied methodologies (influenced in particular by the credibility revolution Angrist & Pischke [2000]) to renew the field by focusing on governance issues in terms of mechanisms and broadly speaking by considering a more micro/inter-individual levels. What is interesting is that the inflexions on the matter can clearly be observed on the 2008-2012 period. This triggered the self-recognition by the scholars in the field of an evolution of its nature that explained to a large extent why a consensus emerged to decide in 2014 to requalify the society with a name reflecting this evolution by dropping the “school of thoughts” aspect of the initial name and by pointing out the increasing focus on “organizational” phenomena beside “institutional” ones.

References

Blei, D. M., & Lafferty, J. D. (2006, June). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning* (pp. 113-120). ACM.

Hu, Y., Boyd-Graber, J., Satinoff, B., & Smith, A. (2014). Interactive topic modelling. *Machine learning*, 95(3), 423-469.

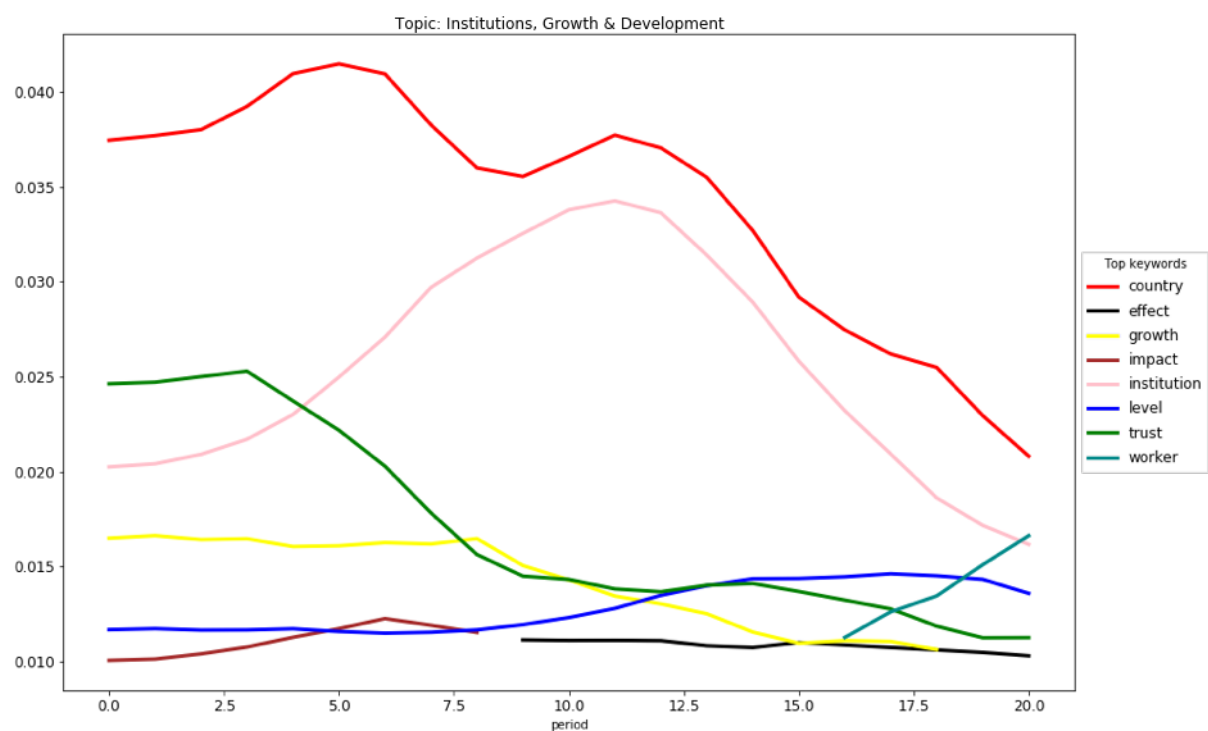
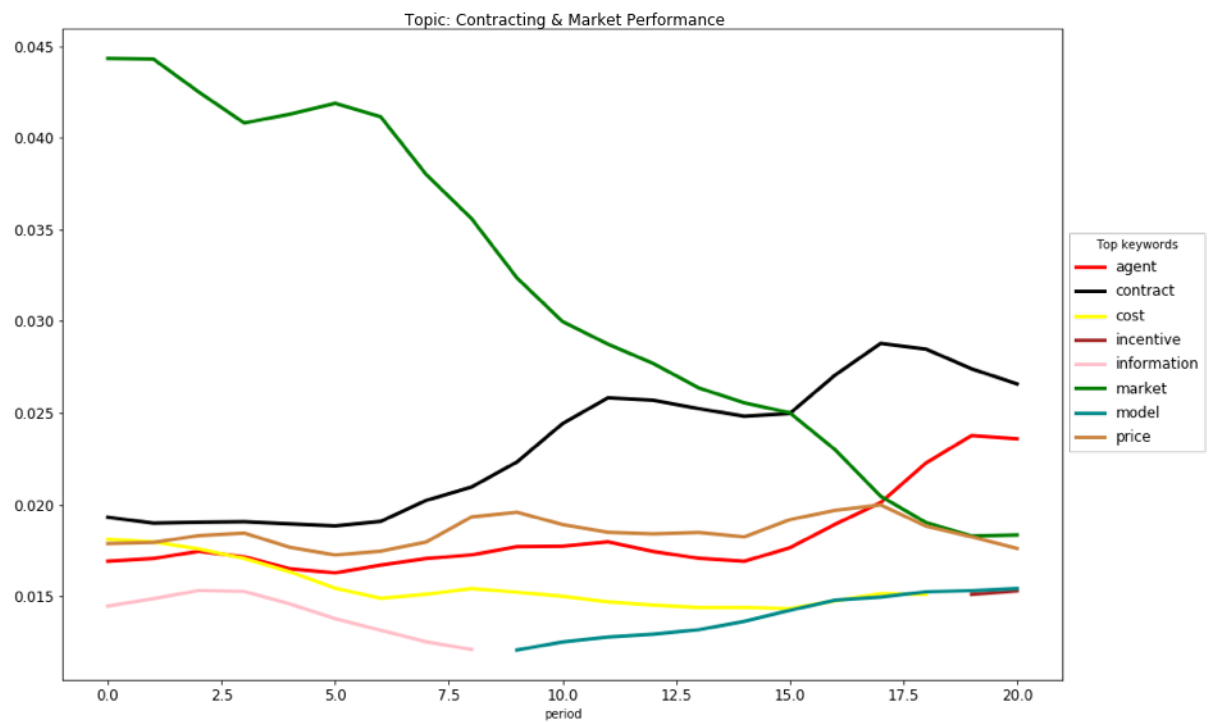
Jacobi, C., van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89-106.

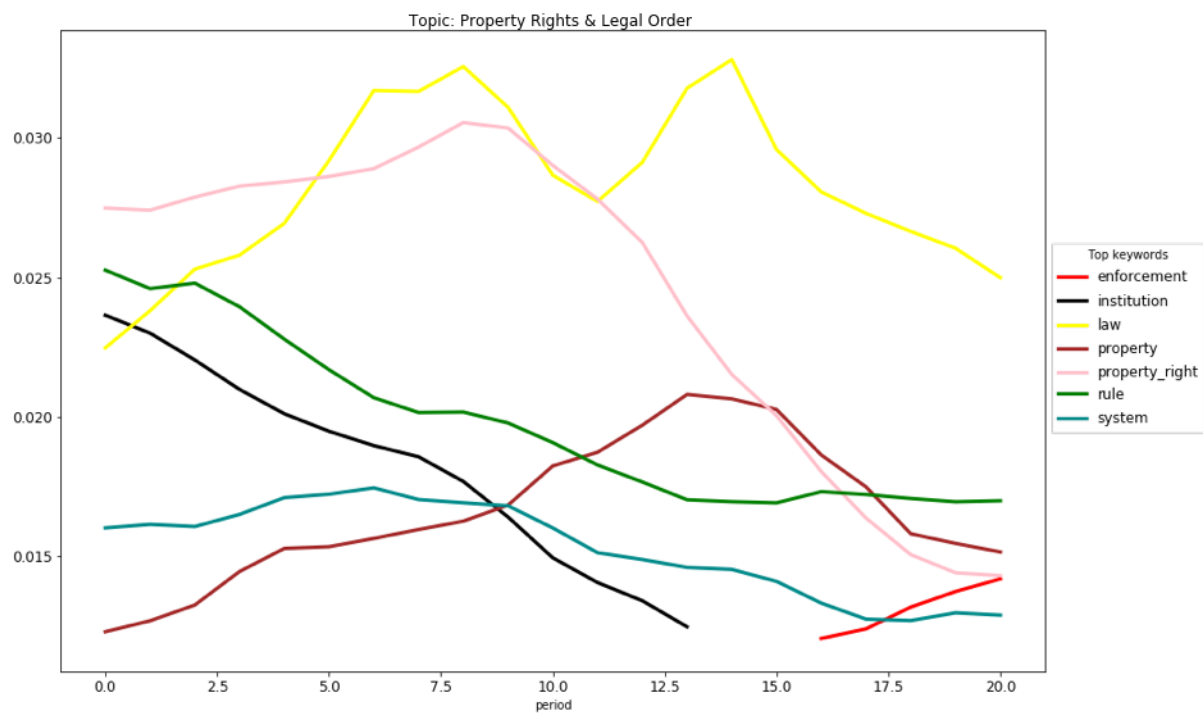
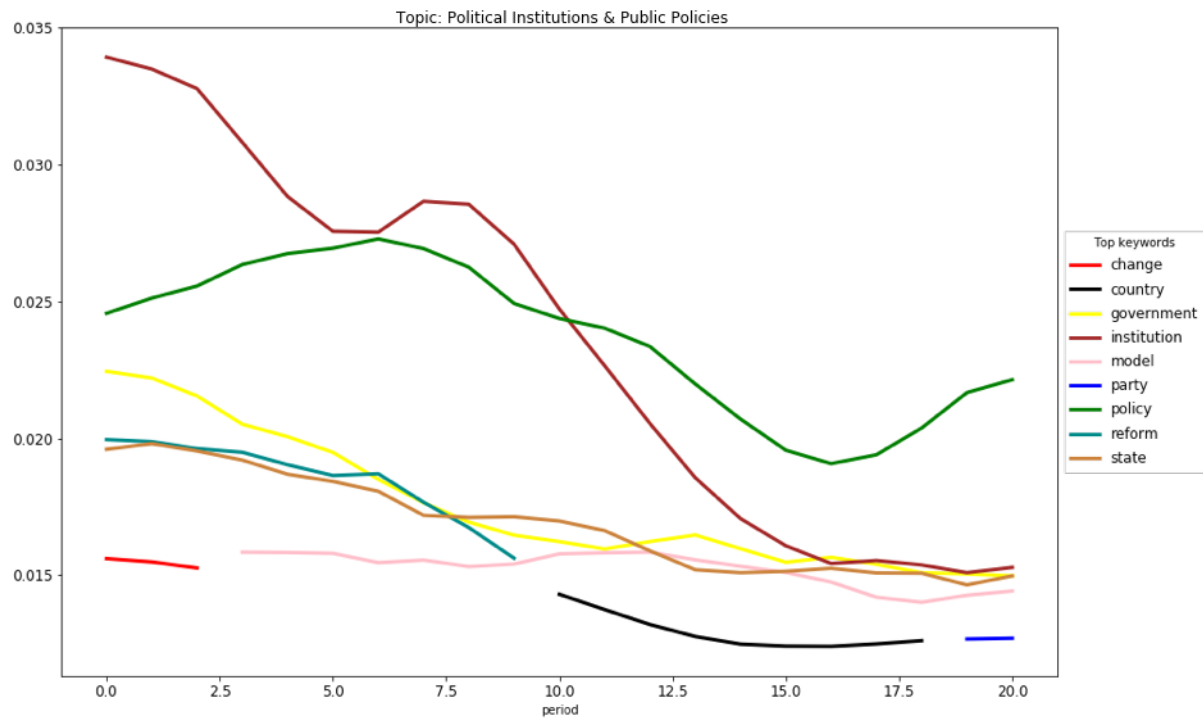
Kaplan, S., & Vakili, K. (2012, July). Identifying breakthroughs: using topic modelling to distinguish the cognitive from the economic. In *Academy of Management Proceedings* (Vol. 2012, No. 1, p. 11036). Briarcliff Manor, NY 10510: Academy of Management.

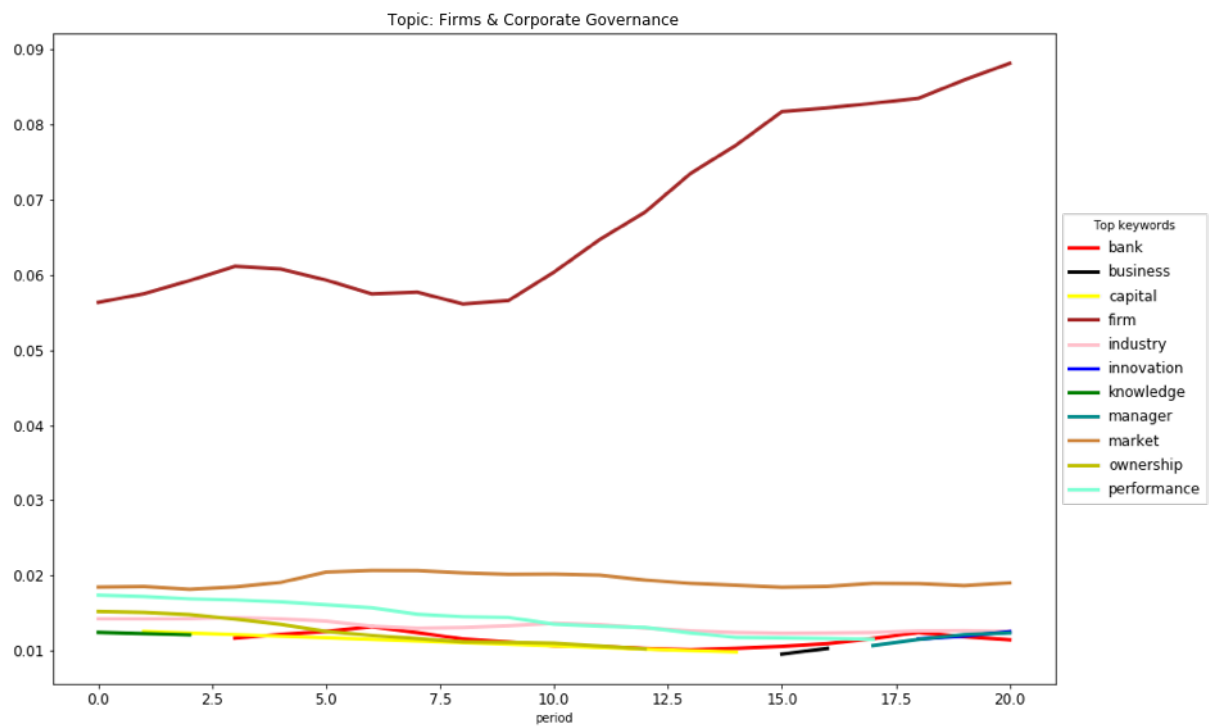
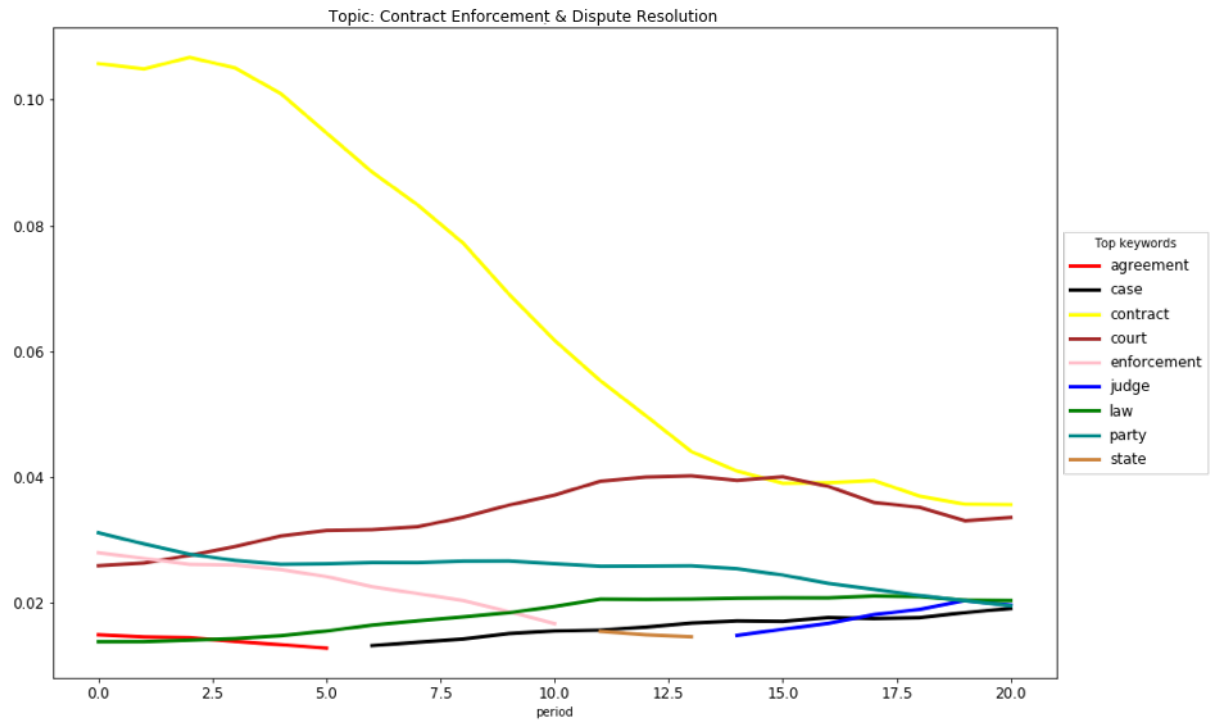
Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408). ACM.

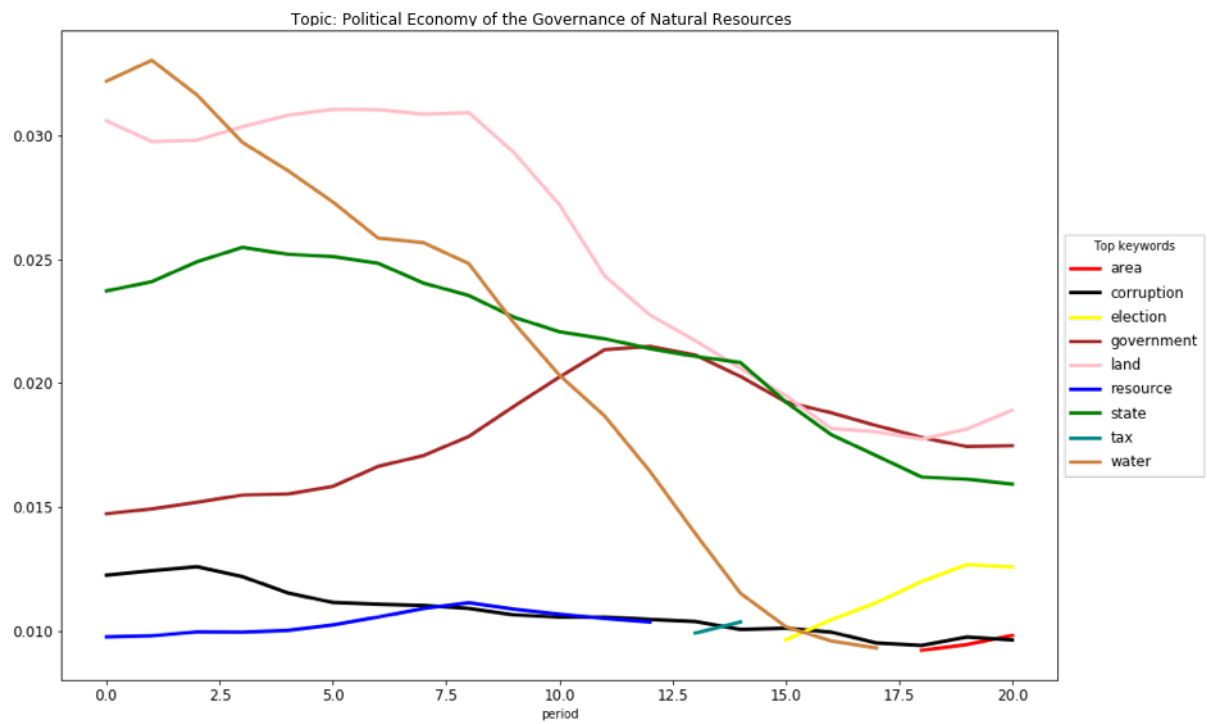
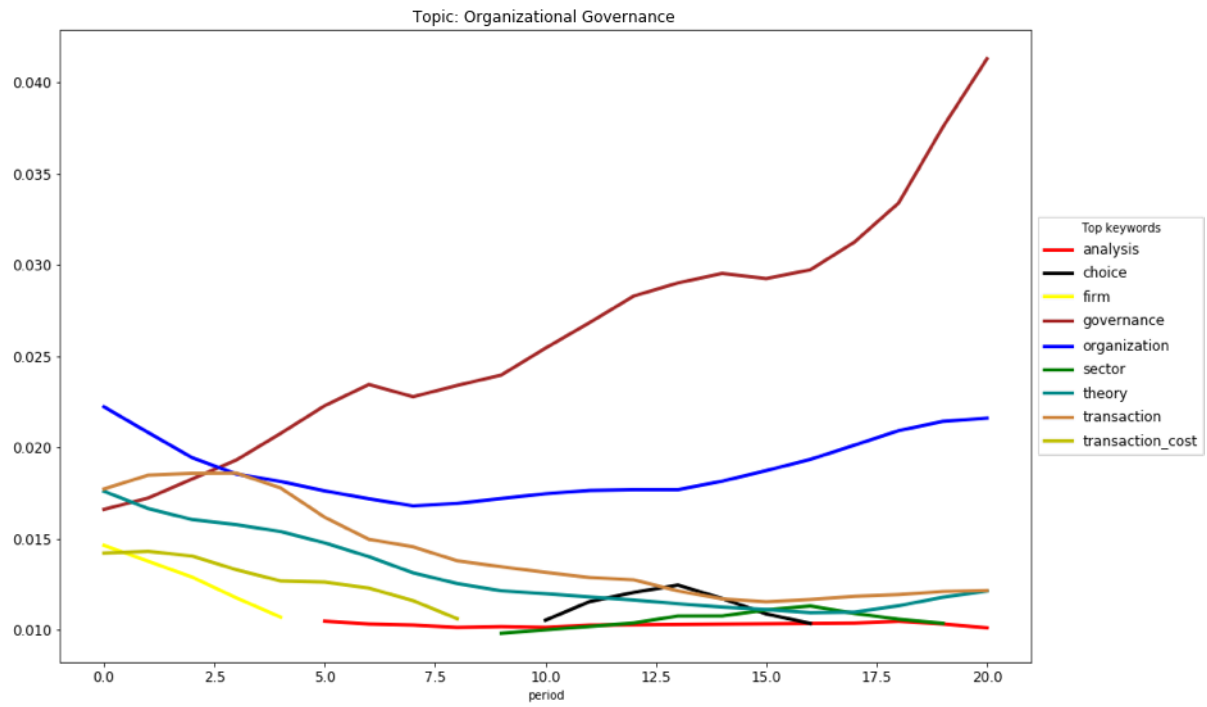
Vulić, I., De Smet, W., Tang, J., & Moens, M. F. (2015). Probabilistic topic modelling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management*, 51(1), 111-147.

Appendix 1 : Dynamic Topic Modelling by 10 most Frequent Words





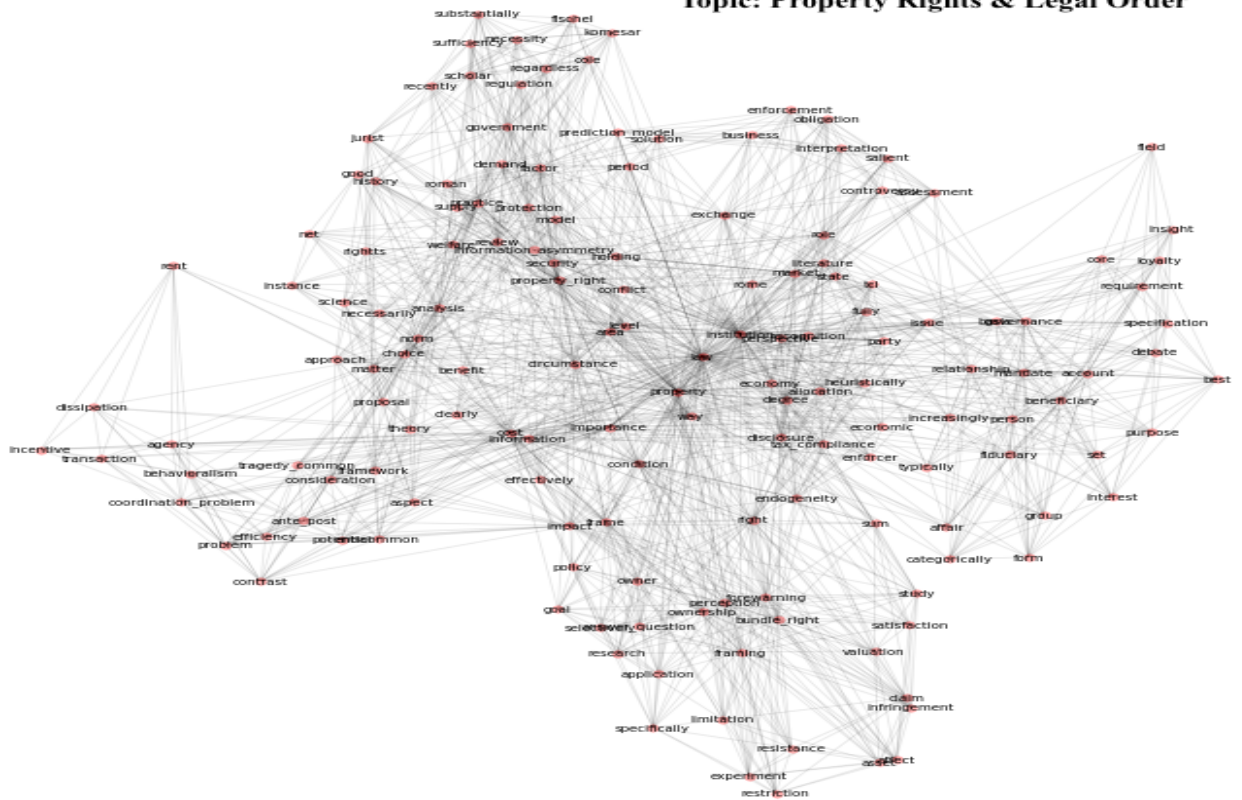




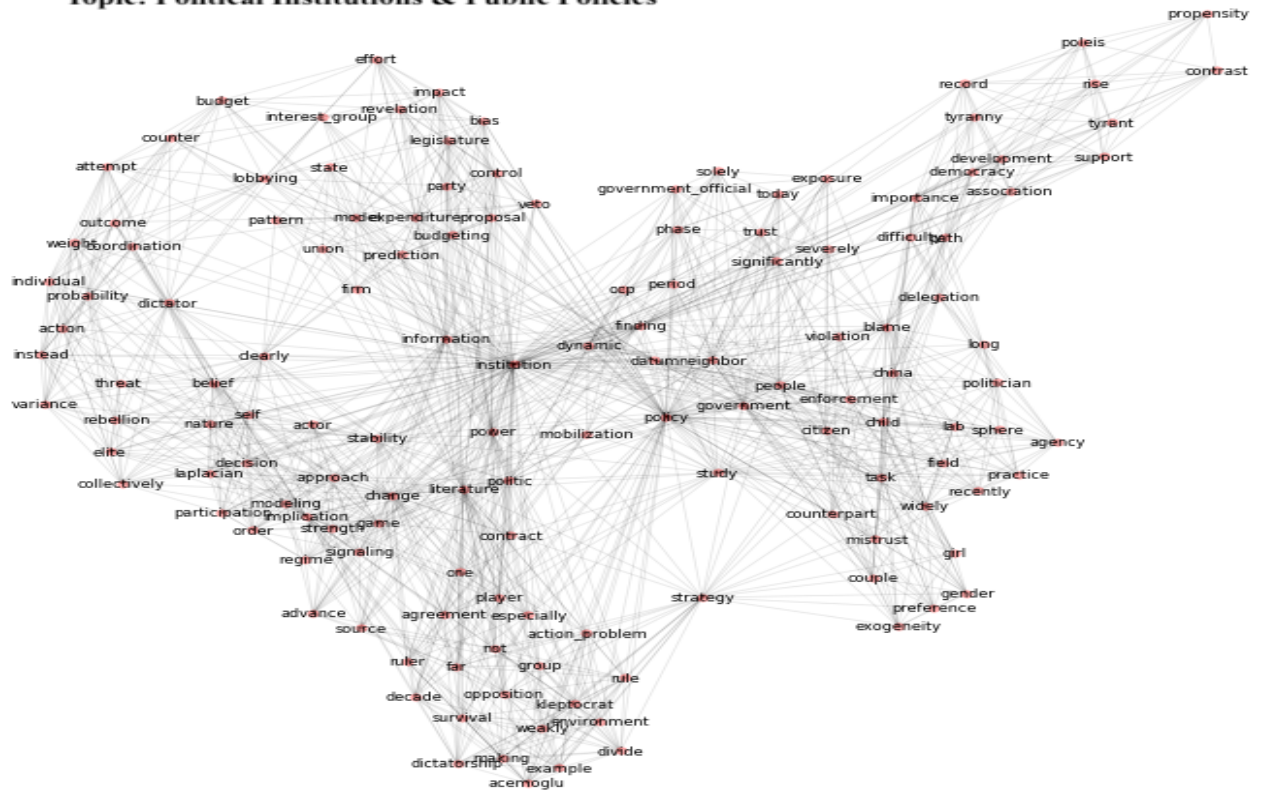
Semantic Networks that represent semantic relations between concepts in the collection of abstracts of corresponding topic.



Topic: Property Rights & Legal Order



Topic: Political Institutions & Public Policies



[illegible]

Appendix 3 :

Automatic text summarization of the collection of abstracts for corresponding topic.

Topic: Contracting & Market Performance

Learning whether there are systematic differences between decisions made by groups and individual agents in market environments characterized by risky outcomes is needed.

Private-to-private contracts are less rigid than private-to-public contracts for the provision of the same service.

The level of contractual rigidity increases with local political contestability for private-to-public contracts.

The incentive contract used to induce the agent's effort can include various incentive instruments, each of them based on one of some performance measures.

There has been little empirical study of the relationships among the incentive instruments in one agent's incentive contract.

A principal-agent model is a model where a risk-neutral principal delegates to a risk-neutral agent the decision of whether to pursue a risky project or a safe one.

Topic: Institutions, Growth & Development

Information asymmetries might provide an answer on how an individual takes a decision to trust an institution based on prior beliefs, current political agents and institutional quality.

Surveys on the role of public and private R&D in several developed countries and the impact of institutional choices on economic growth.

If certain political institutions encourage economic growth via greater investment in physical and human capital, one should observe a change in these institutions to induce a change in firm owners' beliefs about the stability of government policy and the protection of property rights.

There is the effect of differences in informal institutions on economic relations between countries.

Local political institutions, not human capital, were the main factors for the development of the high standard of living in many regions.

The study presents an attempt to quantify institutional changes and examine the respective effects of institutions on the path of long-run economic growth and development for a large panel of countries.

Topic: Political Institutions & Public Policies

Institutional economics has explored the relationship between markets and firms, governance institutions and economic performance, and why societies choose inefficient institutions, among other crucial questions in political economy.

Political institutions affect public policy, the changes in political incentives impact state firms' decisions.

There are strong empirical and theoretical grounds for believing that inefficient policies and institutions are prevalent, and that they are chosen because they serve the interests of politicians or social groups holding political power, at the expense of the society at large.

The political economy literature suggests that redistributive policies are subject to political institutions such as electoral rules, political competition as well as the degree of fiscal and political decentralization in the country.

We consider the allocation of authority of policy making between politicians and bureaucrats and analyze the efficiency of the political governances in several political-administrative institutions.

There is some evidence for a non-monotonic relationship between the level of political accountability and the size of government, which could be explained by the dominance of different incentive mechanisms between politicians and bureaucrats.

Topic: Property Rights & Legal Order

Evidence suggests that at the macro-level property rights have a greater influence on long-run economic performance than contracting institutions.

Since property rights incompleteness is a pervasive feature of every bundle of property rights, both property rights and externalities co-evolve over time as a result of a complex dynamics between public and private institutions which involves changes in technology, preferences, information, knowledge, wealth and institutional setting.

In property rights theory, firm is an organizational response to reduce transaction cost associated with hold-up of using market mechanism.

Using the paradox of the relative stability of urban form and illegality of property rights held by many individuals as an analytical entry point, the study uses tools of new institutional economic theory to develop a conceptual

framework linking informal property markets and the production of the built environment in developing countries.

While there is broad agreement that the “rule of law” – a stable and predictable process by which laws are implemented, enforced, and changed – is important to economic growth, the nature of the circumstances under which a rule of law arises, and why some states are able to achieve an effective rule of law while others are not, remain open questions.

The analyses tend to overstate the role of “private ordering” and disregard the two key elements of property law: first, the essential conflict between property (that is, in rem) enforcement and transaction costs; and, second, the institutional solutions created to overcome it, mainly contractual registries capable of making truly impersonal (that is, asset-based) trade viable when previous relevant transactions on the same assets are not verifiable by judges.

Topic: Contract Enforcement & Dispute Resolution

One sketches two models of governance of the cooperative firms: the first one describes the mechanisms of contractual governance, while the second is concentrated on the «voluntary and collective» norms and enforcement mechanisms at the level of professions and industries.

The empirical analysis strongly supports the models: (1) the farmers’ responses can be explained by cost-benefit calculations regarding the use of courts, (2) the costs of court enforcement are significantly affected by the attributes of the contract and the value of the relationship, (3) indirect costs of court enforcement, such as the disruption of a valuable relationship, play a significant role in explaining the threshold of taking legal action, and (4) the threshold of court enforcement is affected by the availability of alternative enforcement mechanism.

One seeks to understand the variation of the costs of enforcing contracts by examining how informal cultural enforcement mechanisms may substitute or complement formal legal procedures ultimately shaping the costs of contract enforcement.

The results suggest that variation in contract enforcement costs may be better explained from cross-country differences in culture than formal legal rules.

One explores the role of the courts and of the law in the economic theory of incomplete contracts. court efficiency on economic growth in a within-country setting.

Based on intermediation theory and the insights from new institutional economics, three key factors responsible for intermediation in foreign trade are identified: risk associated with the non-enforcement of contract or property rights; transaction cost reductions through the exploitation of economies of scale from trading; reinforcement of trust in a network of intermediaries.

Topic 5: Firms & Corporate Governance

The measures of management quality help observe large variation in average management practice levels across ownership types, with low scores for domestic private firms and high scores for state-owned firms, after controlling for firm size, labor skill level, and industry.

Cooperation among firms and between firms and workers allows training firms to replace a quitting worker with an equivalent from the labor market at low cost.

The results show that government decentralization – “federalism” – plays an important role in improving the performance of not just collective firms, but also state-owned and mixed public/private ownership firms, in theory, however, smaller firms may also have advantages over larger firms. The study examines how political, institutional, and economic factors are related to a country’s decision to privatize state-owned banks. Four main findings are that (1) the wage policies of worker-managed firms are more egalitarian than those of conventional firms; (2) in worker-managed firms, high-ability members are more likely than other members to exit; (3) the hazard ratio of high-ability members is lower for founding members and for those employed by worker-managed firms in which there is less pay compression; and (4) high-ability members are less likely to quit when labor market conditions in the capitalist sector are less attractive. Some important financial connections: (1) relationship banking as a close relationship between a firm and a bank, resulting from long-term lending with inside information, (2) relationship investing as a close relationship between a firm and a non-bank institutional investor, where direct control is exerted via large shareholdings or inside equity, (3) transaction finance by bonds or stocks or by arm’s length finance through intermediaries.

Topic: Organizational Governance

The concept of firms as a ‘nexus of contracts’ between members is well established, however as yet there is little research that explores the internal governance choices made by firms using the transaction cost framework.

The effects of the corporate governance system on the relationship between ownership structure, organizational behavior and firm performance are examined.

Inside a production inadequate institutional environment resulting in a high probability for a firm to be faced with adverse external conditions resulting in the high transaction cost brings about the lack of incentives to invest into energy saving.

Two main consequences derive from the theoretical redefinition of the interaction space of economic agents: i) enrichment of the idea of “transaction cost”, with the use of the idea of the existence of “costs of using the price mechanism”; and ii) the effective integration of the concept of embeddedness to the perspective of governance. The dynamic of the building of the governance systems that shape the (global) digital infrastructure and the information society, can largely be analyzed as the result of a process by which the US Federal Government and a network of others, hand-in-hand with a set of firms and non-state actors, played a strategy favoring the diffusion of technologies and economic models ensuring an enduring leadership.

Corporate governance codes are an increasingly prominent feature of the regulatory landscape in many countries, yet remarkably little is known about the determinants of corporate governance reform.

Topic: Political Economy of the Governance of Natural Resources

Disentangling the land rights within family groups sheds light on the debate regarding the individualization of land rights; on the fact that the household constitutes, in such a context, a weak analytical unit; on the potential impact of a land titling policy; on the functioning of the land market (through understanding the formation of supply); on the productive use of the land; and on land conflicts.

Drawing from transaction cost economics and property rights literature we contend that firms with unique resources will use a different alignment between transaction attributes (specific investments) and governance structures (contracts versus decision control) to generate different kinds of benefits.

The legal-institutional underpinnings of the market for urban land use rights are examined, taking German public and private land use law as a case in point.

Until the end of a multi-century process, an elite dominant coalition, including the military, produce stability by denying non-elite forces access to the political process and dividing the monopoly “rents” among themselves.

To illustrate the nature of the socio-economic and political equilibrium one analyzes: 1) the failure of land reform as a strategy allowing for the formation of assets among poor peasants; 2) the limited role played by stock markets in the formation of capital for the vast majority of firms; 3) the regulatory framework that has assured the dominance of a small number of economic organizations, including monopolistic firms and unions.

Economic policies, levels of corruption and rule of law in new or poor democracies are frequently no better than in non-democratic countries.