In the traditional view, as developed for hard and mature sciences, the evolution of scientific paradigm consists in "*the successive transition from one paradigm to another via revolution*" [Kuhn, 1970, p.12]. However, a scientific field is usually composed by several research paradigms either competing or addressing different issues, and a revolution in one of those necessarily involves effects and readjustments in the entire discipline. Moreover, each new paradigm carries the legacy of the existing knowledge of past paradigms, which is often recombined into the new one. This is especially true for social sciences, in which the identification of clear scientific paradigms in the sense of Kuhn is often blurred and it is probably more correct referring to "research traditions" [Laudan, 1978].

However, whether you call paradigms or traditions, the existence of patterns of thoughts which are legitimate contributions to a theory is undeniable. Thus, we can postulate that the evolution of knowledge in a scientific field is generated among a community of researchers which share a semantic area to define specific research issues, describe methodologies, and lay down results. Thus, the heterogeneity of the research tradition of a scientific field can be described with semantic analysis. The idea that some measure of words co-occurrence reveals an underlying epistemic pattern and, therefore, it can capture the essence of evolution in science is not a new one. Despite the difficulty in programming, the first attempts date back to the work of Callon et al. [1983] and refined when the first open code have been made available a decade later [Vlieger and Leydesdorff, 2011, Leydesdorff and Welbers, 2011].

The challenge of classifying science on the basis of its semantic content has found a renewal with the diffusion of machine learning techniques and, in particular, in the subfield of unsupervised learning [Leydesdorff and Nerghes, 2015]. Topic modeling includes a family of algorithms [Blei et al., 2003], which are particularly performant in extracting information from large corpora of textual data by reducing dimensionality. This feature has been clearly recognised in mapping science [Suominen and Toivanen, 2015] or news [DiMaggio et al., 2013]. Alghamdi and Alfalqi [2015] review four major methods of topic modeling, including Latent Semantic Analysis (LSA), Probabilistic LSA, Latent Dirichelet Allocation (LDA) and Correlated Topic Model (CTM). The LDA proposed in [Blei et al., 2003] is one of the most diffused approaches. LDA retrieves latent patterns in texts on the basis of a probabilistic Bayesian model, where each document is a mixture of latent topics described by a multinomial distribution of words. One of the major limitations of LDA lies on its inability to model and represent relationships among topics over time [Alghamdi and Alfalqi, 2015].

In this paper, we address a major recurring issue in topic modeling, that is the topic dynamics, or, in other words, we test a method to track the transformation of topics over time. As stated by Blei and Lafferty [2006], LDA is a powerful approach to reduce dimensionality, but it assumes that documents in a corpus are exchangeable. On the contrary, articles and themes are sequentially organized and evolve over time. Therefore, it is not only relevant to develop a statistical model to determine the evolving topics from a corpus of a sequential collection of documents, but also to measure and describe the transformation of topics and their appearance and disappearance.

In the literature of information retrieval, the dynamics of topics has been faced with two approaches [He et al., 2009]: a discriminative one monitors a change in the distribution of words or in the mixture over documents, while a generative approach searches for general topics over the whole corpus and, then, it assigns the documents which belong to each topic [Bolelli et al., 2009, He et al., 2009].

Specifically Blei and Lafferty [2006] introduced Dynamic Topic Modeling (DTM), a class of generative models in which the per document topic distribution and per topic word distributions are generated from the same distributions in a previous time frame. This approach has been very influential since it imposes a connection between the sets of topics at different periods and allows

to track the evolution of a single topic over time.

DTM performs very well in capturing the evolution of a single topic. However, the evolution of knowledge is much more complicated that the change of relative importance of words within a topic, since it may involve also the creation of new topics, their mutual re-combinations and, eventually their possible demise. The major contribution of the paper is the conceptualization and formalization of the evolution of knowledge, conceived as different streams of semantic content which continuously appears and disappears, merges and splits. Thereby we propose an original method based on inter-temporal bimodal networks of topics compute the key elements in the evolution of knowledge.

Moreover, the ultimate goal of the paper is not to track in detail what happens within a single topic, but rather to develop indexes which can measure at the aggregate level some properties of the observed knowledge dynamics, such as an overall degree of novelty or the level of turbulence at specific time windows.

The paper is organized as follows: in the next section, we suggest a method to analytically conceptualize and measure different patterns of topics evolution. Section 2.2 translates it into an algorithm which calculate some measures of merging, splitting and novelty of the topics generated by the LDA. In section 3.1, a simple simulation tests the robustness of the method on artificial data. Finally, in Section 4, the same algorithm is applied to a large dataset of papers in economics: main results are presented and discussed by describing the evolution of the topics in the economic science in the past century.

2 A Conceptualization of Knowledge Evolution

In this paper, we focus on the dynamic evolution of topics over time. With DTM, each topic K_t is linked to K_{t+1} creating a topics chain which spans the years covered by the documents. Specifically, Blei and Lafferty [2006] maps each topic at time t-1 into a topic in t by chaining the per document topic distribution α_t and the per topic word distribution, $\beta_{t,k}$ in a sate space model with a Gaussian noise:

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

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

This approach is highly performing to track incremental changes of the same topic but it does not focus on revealing neither birth nor death nor possible combinations of topics and it imposes a constant number of topics within the model. On the contrary, we are interested to discover the structural change of topics in a corpus and to understand the underlying topic dynamics which explain it. Thereby, we do not focus on the evolution of the single topic. The inter-temporal link across topics is not a constraint in the estimation of the model as in the DTM, but it is introduced ex-post in the empirical analysis by looking at the similarities (co-occurrence of words) amongst topics generated by independent LDAs. More in detail, while DTM models sequences of compositional random variables by chaining Gaussian distributions (thus directly embodying topics dynamics in the model), our approach operates on single and static LDAs in order to track and measure such dynamics out of the model.

The evolution of a topic structure of a corpus accumulating knowledge overtime takes place because of two main reasons. On the one hand, any epistemic community (say for instance journalists

or scientists) can shift their intellectual interest to new issues and problems, which will result in different choices, frequencies and co-occurrence of words. On the other hand, language is subject to a constant evolution, in which new words, named entities, acronyms, etc. appear while other ones disappear due to an increasingly lesser use of them by the same community. We rule out this second scenario, by assuming that in the short time frame the language is fairly stable.

Under this assumption, when comparing the topics generated by a topic modeling exercise in two different, although adjacent, time windows, we should be able to capture the evolution of the scientific debate and highlight the birth, death and recombination of topics. On the one extreme, we can find a situation in which knowledge does not evolve and thus topics are stable. On the other, we figure out the maximum of turbulence in which new topics emerge without any semantic relation with the incumbent ones. In the latter case, we may assume the death of past topics and the birth of new ones. In between the two ideal cases, we can also draw a continuum in which we can observe both deaths and births of topics. Finally, in a most interesting scenario, rather than observing stability or turbulence, knowledge may evolve recombining existing topics in both old and new ones. Table 1 summarizes five typical patterns of knowledge evolution and their interpretation within a topic modeling framework.

Table 1: Topic modeling and typical patterns of knowledge evolution

Stability	a topic A exists at time t and t+1
Birth	the topic A at time t+1 has no antecedent at time t
Death	the topic A at time t disappears at time t+1
Merging	multiple topics at time t combine in a new topic A at time t+1
Splitting	multiple topics at time t+1 share an antecedent at time t

Figure 1 presents the five ideal types of knowledge evolution as a proximity network of topics, that we mathematically formalize as follows. Let us consider M topics emerged as the result of a topic modeling exercise from a corpus of articles at time t and N topics at time $t + 1$. We tackle the critical problem of tracking the transformation of the set of topics $M = (1, \ldots, A, \ldots, M)$ at t into the set of topics $N = (1, \ldots, a, \ldots, N)$ at $t + 1$. Specifically, we are interested in measuring the magnitude of the various phenomena such as birth, death, merging, and splitting. Consider a similarity index based on word co-occurrence, $simil$[1], between each couple of topics (A, a) with $A \in M$ and $a \in N$ and consider the similarity matrix S $(M \times N)$
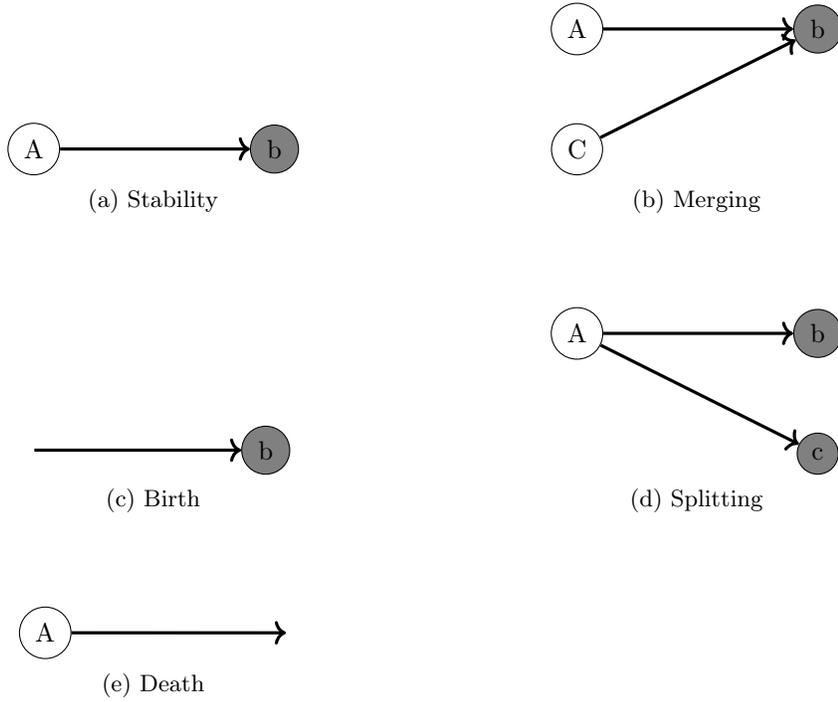
$$
\text{S} = \begin{array}{c} A \\ \vdots \\ M \end{array} \begin{matrix} a & \quad \ldots & \quad N \\ \left[\begin{matrix} simil_{1,1} & \ldots & simil_{1,N} \\ & \ddots & \\ simil_{M,1} & \ldots & simil_{M,N} \end{matrix} \right] \end{matrix}
$$

For the sake of clarity and with reference to Figure 1, let us consider the minimal example in which $M = (A, B)$ and $N = (a, b)$

---

[1]Typically, this index is the cosine similarity index, as it is used in the empirical part of the paper

Figure 1: Ideal types of topic evolution

(a) Stability

(b) Merging

(c) Birth

(d) Splitting

(e) Death

$$\text{S} = \begin{array}{c} \\ A \\ B \end{array} \begin{array}{cc} a & b \\ \left[\begin{array}{cc} \alpha & \beta \\ \gamma & \delta \end{array} \right] \end{array}$$

The network representation allows to visualize the five ideal types of knowledge evolution: Table 2 summarizes them and the necessary and sufficient conditions on the values of the similarity index to observe such cases. However, with a higher number of topics, a derivation of the conditions on the values of the similarity index would be cumbersome. Moreover, Table 2 depicts ideal situations only, while the observed reality usually deals with a continuous mixture of the paradigmatic cases presented above. For instance, already in the case with $M = 4$ and $N = 3$ depicted in Figure 2, the analysis becomes strenuous.

With this purpose in mind, we consider the similarity matrix S as the incidence matrix of M over N. We can thus employ S to create a bi-adjacency matrix D, and thus consider Figure 2 as the resulting bipartite network in which M and N are the sets of nodes, while the elements of the matrix are the weights of the edges.

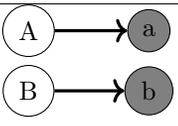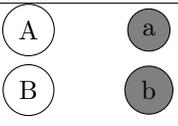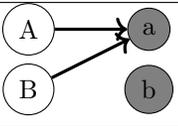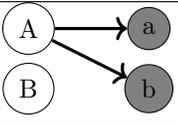Table 2: Bimodal network and empirical indexes

Network	Matrix param.	Cases
A → a B → b	$\alpha, \delta \neq 0 \ \gamma, \beta = 0$	STABILITY: no births, no deaths
A a B b	$\alpha, \delta, \gamma, \beta = 0$	INSTABILITY: births and deaths
A → a B → a	$\alpha, \gamma \neq 0 \ \delta, \beta = 0$	MERGING: no deaths, but births
A → a A → b B	$\alpha, \beta \neq 0 \ \delta, \gamma = 0$	SPLITTING: no births, but deaths

Figure 2: bipartite network of topics of two time windows
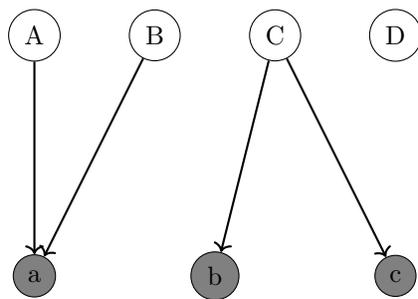
$$D = \left[\begin{array}{c|c} 0 & S \\ \hline S^T & 0 \end{array}\right] =$$

$$= \begin{array}{c} \\ A \\ B \\ ... \\ M \\ a \\ b \\ ... \\ N \end{array} \begin{array}{cc} \begin{array}{ccccccc} A & ... & M & a & b & ... & N \end{array} \\ \left[\begin{array}{ccc|cccc} 0 & 0 & 0 & & & & \\ 0 & 0 & 0 & & & S & \\ 0 & 0 & 0 & & & & \\ 0 & 0 & 0 & & & & \\ \hline & & & 0 & 0 & 0 & 0 \\ & & & 0 & 0 & 0 & 0 \\ & S^T & & 0 & 0 & 0 & 0 \\ & & & 0 & 0 & 0 & 0 \end{array}\right] \end{array}$$

We now show how this representation can help measure the magnitude of births, deaths, merging and splitting.

Births and deaths can be easily calculated from the matrix S. A row sum equal to zero highlights a death, while a column sum equals to zero indicates a birth. A death means that the semantic legacy completely disappears while a birth means that a topic carries no semantic similarity with other topics in the past. Once again it is important to notice that these cases are extreme scenarios while in the reality we observe a continuum between births and deaths. We might thus calculate an index $Novelty\_i\,(NI)$ for each topic i at time $t+1$ where for $NI\_i = MAX$ we have a birth, that is a topic with no similarity to any other previous one. For higher value we have a higher novelty of the topic. We can also measure an average change in NI on the overall structure of a scientific field by looking at distributions of these indexes over the topics. For instance, let us consider the *Novelty Index* and the average, defining:

$$NI_j = 1 - \frac{\sum_i^M S_{i,j}}{M} \tag{3}$$

where j is the index of the j-th column in the matrix S, and

$$NI = 1 - \frac{\sum_i^M \sum_j^N S_{i,j}}{M * N} \tag{4}$$

We take the average of all the cell values in matrix S. If the similarity index is bounded between 0 and 1, such it is the very common case of the cosine similarity index, thus NI ranges from 0 to 1. For very small value of novelty, new topics show different word distribution from old ones.

As mentioned, transformation of topics can take the form of merging and splitting. We say that a merging occurs if a topic at time $t+1$ shows a high similarity with two topics at time t, meaning that the semantic universe of A and B at t (as in figure 2) is combined in the topic a. Similarly, we can say that a split occurs if the semantic legacy of one topic at t is to be found in multiple topics at $t+1$ as in the case for topic C.

To analyse the intensity of a merging we can project the bipartite network of figure 2 into its two 1-mode-network of figure 3. This is achieved by a matrix multiplication $S \times S^T$ for the merging and $S^T \times S$ for the splitting which result in two matrices $P^{merging}$ and $P^{splitting}$ of dimension respectively $M \times M$ and $N \times N$. Please note, that for the properties of matrix multiplication

Figure 3: 1 mode network, projection of Figure 2
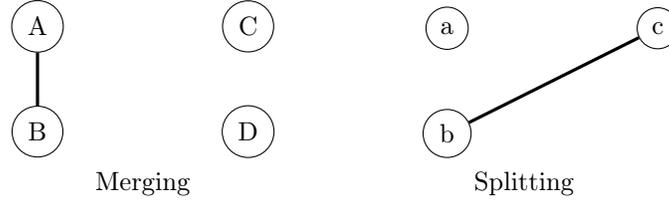
Merging   Splitting

$P^{merging}$ and $P^{splitting}$ are always square matrices, even when the number of topics in two periods differs.

The network is represented by the matrix P

$$P^{merging} = \begin{array}{c} \\ A \\ B \\ ... \\ M \end{array} \begin{array}{cccc} A & B & ... & M \end{array} \left[\begin{array}{cccc} & & & \\ & S \times S^T & & \\ & & & \\ & & & \end{array} \right]$$

$$P^{splitting} = \begin{array}{c} \\ a \\ b \\ ... \\ N \end{array} \begin{array}{cccc} a & b & ... & N \end{array} \left[\begin{array}{cccc} & & & \\ & S^T \times S & & \\ & & & \\ & & & \end{array} \right]$$

The matrix transformation allows us to draw the 1-mode-network as in Figure 3, which represents the merging and splitting between two time windows. The matrix formulation of the network is also useful for computing the intensity of merging and splitting on the basis of the two relative matrices P. Let us consider the matrix $P^{merging}$ in a minimal example of the table 2

$$P^{merging} = S \times S^T = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \times \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} = \begin{bmatrix} \alpha \cdot \alpha + \beta \cdot \beta & \alpha \cdot \gamma + \beta \cdot \gamma \\ \alpha \cdot \gamma + \beta \cdot \delta & \gamma \cdot \gamma + \delta \cdot \delta \end{bmatrix} \qquad (5)$$

The matrix P is always symmetric and, for our purpose, we focus on the low triangle. The merging is captured by the number outside the diagonal $(\alpha \cdot \gamma + \beta \cdot \delta)$, where $(\alpha \cdot \gamma)$ is the intensity of the merging of A and B in a, while $(\beta \cdot \delta)$ is the intensity of the merging of A and B in b. In this exemplary case shown in Table 1, β and δ are equal to zero and α and γ are different from zero: thus, we have a merging between A and B as depicted in Figure 3.

Mutatis mutandis, we can consider the case of splitting. Once again, the low triangle off the diagonal highlights the intensity of split with $(\alpha \cdot \beta)$ the split of A in a and b, while $(\gamma \cdot \delta)$ the split of B.

$$P^{split} = S^T \times S = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \times \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} \alpha \cdot \alpha + \gamma \cdot \gamma & \alpha \cdot \beta + \delta \cdot \gamma \\ \alpha \cdot \beta + \gamma \cdot \delta & \beta \cdot \beta + \delta \cdot \delta \end{bmatrix} \qquad (6)$$

When we have a large number of topics in both time windows, we can use this formulation to create indexes measuring the intensity of merging and splitting or other properties of the transition. Specifically, we aim at comparing the values below the diagonal with those on the diagonal. We

Figure 4: Similarity network among topics in t
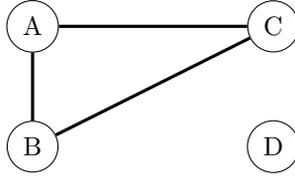
thus create a normalized matrix in which all elements of the diagonal and below the diagonal add up to one.

$$P^{merging}_{normalized} = P^{merging} \cdot \frac{1}{\sum_{i \leq j} P(i,j)} \tag{7}$$

In this way, we can compute a $MergingIndex$ (MI) which takes value 0 when no merging occurs and it ranges up to an upper limit which can not exceed 1.

$$MI = 1 - trace(P^{merging}_{normalized}) \tag{8}$$

Symmetrically, we calculate a $SplittingIndex$ (SI)

2.1 Conditional dependence

A last important issue to be addressed consists of the impact of the conditional dependence of topics at time t and its relation with the 1-mode network projection. Two topics at t can appear to merge into a topic at $t+1$ only because they are already similar to each other at time t. In this case we might run the risk of identify a spurious process of merging. However, it is possible to account for this dynamic conditional dependence. We can compute a similarity index among topics at time t, $simT$, which can also be represented by a network.

$$Q = \begin{bmatrix} simT_{1,1} & \ldots & simT_{1,M} \\ & \ddots & \\ simT_{M,1} & \ldots & simT_{M,M} \end{bmatrix}$$

Note that Q is a symmetric matrix, with the same dimension ($M \times M$) as $P^{merging}$.

The same procedure can be applied to topics at $t+1$. In this case, we obtain a matrix ($N \times N$), with the same dimension of P^{split}.

In order to take into account the conditional dependence, we might consider $R^{merging,splitting} = (P^{merging,splitting}|Q^{merging,splitting})$ and recompute the indexes, substituting R with P. There exist different ways to operationalize the dependence. Probably the most sophisticated one would be to encode the overall conditional dependence structure within a graphical network [Jordan, 1998, Lauritzen, 1996]. However, we might also consider that the similarity measure has a scalar meaning which goes beyond a simple probabilistic relation. For this reason, we surmise that the conditional dependence can be at best considered by dividing or subtracting element by element the two matrices: in the developed algorithm (see next paragraphs), we divide. Table 3 summarizes the indexes we use and their range.

Table 3: Measuring change in topic modeling

Type of change	Index	Min	Max
Introduction of new semantic areas or legacy from the past	NI	0	1
Merging of the semantic content of topics	MI	0	1
Splitting of the semantic content of topics	SI	0	1

2.2 The Proposed Algorithm

This paragraph describes the algorithm which we developed to operationalize the former theoretical approach. Our example relies on the Latent Dirichlet Allocation (LDA) [Blei et al., 2003], although this methodology does not involve any assumption in the way topics are created. LDA is a generative model that summarizes the documents through a mixture of topics, where each topic is a probability distribution in the dictionary. The algorithm first generates a database which allows query of documents per time period. Thereafter, it divides the dataset into unigrams where stopwords are eliminated according to the NLTK list (www.nltk.org). Finally, we have applied the *Porter Stemmer* [Porter, 1980] on individual words. This algorithm transforms (or truncates) every word in a morphological root form. We create a subset per each T time window and compute N_t topics using standard LDAs [2]. On the generated output we are able to compute the three indexes. For the similarity computation, we use the probability of the first 100 topic's words to generate the vector weights.

Algorithm 1 shows the pseudo-code to compute the time window from t to $t+1$. It simply takes in input the cleaned documents of the selected windows and the number of topics at time t and $t+1$ and returns the merging, splitting and value indexes. In details, the algorithm generate a LDA model for each time window t and $t+1$ and computes the similarity between topics at time t and $t+1$ (and themselves). Then, it computes the matrices $P_{merging}$, $P_{splitting}$ using the similarity matrix S and the matrix Q. The two P matrices are used to compute MI and SI, while the matrix Q is used to compute NI.

3 Evaluation

As to evaluate this approach, we cannot benchmark it with other dynamic methods such as DTM, since we do not track the single topics over time, but we compare adjacent time windows to measure the degree of topics recombination. Therefore, we test the methodology by applying the algorithm on an artificially-generated dataset with controlled characteristics.

3.1 Artificial Data Creation

To generate the experimental datasets, we create artificial topics reflecting natural and realistic textual content. Instead of directly producing topics as sets of artificially-built sets of words, we started from *concept seeds*, used as query of real textual data. A concept seed is a word (or compound word) that represents a concept in a text-based resource. For example, the concept seed *physics* within the *Wikipedia* resource is the *Wikipedia* page about *Physics*. From a set of concept

---

[2] https://radimrehurek.com/gensim/

Algorithm 1 computeSingleWindow(documentSet, $numTopic_t$, $numTopic_{t+1}$)

1: $topic_t \leftarrow \text{LDA}(\text{documentSet}, numTopic_t)$
2: $S_t \leftarrow \text{computeTopicSimilarity}(topic_t, topic_t)$
3: $topic_{t+1} \leftarrow \text{LDA}(\text{documentSet}, numTopic_{t+1})$
4: $S_{t+1} \leftarrow \text{computeTopicSimilarity}(topic_{t+1}, topic_{t+1})$
5: $Q \leftarrow \text{computeTopicSimilarity}(topic_t, topic_{t+1})$
6: $R_{merger} \leftarrow S_t * S_t^T$
7: $R_{split} \leftarrow S_{t+1} * S_{t+1}^T$
8: $Q_{merger} \leftarrow Q * Q^T$
9: $Q_{split} \leftarrow Q^T * Q$
10: $P_{merger} \leftarrow \text{zeros}(R_{merger}.\text{numRow}(), R_{merger}.\text{numCol}())$
11: $P_{split} \leftarrow \text{zeros}(P_{split}.\text{numRow}(), P_{split}.\text{numCol}())$
12: **for** i $\leftarrow 1..R_{merger}.\text{numRow}()$ **do**
13:     **for** j $\leftarrow 1..R_{merger}.\text{numCol}()$ **do**
14:         $P_{merger}[i,j] \leftarrow \frac{R_{merger}[i,j]}{Q_{merger}[i,j]}$
15:     **end for**
16: **end for**
17: **for** i $\leftarrow 1..R_{split}.\text{numRow}()$ **do**
18:     **for** j $\leftarrow 1..R_{split}.\text{numCol}()$ **do**
19:         $P_{split}[i,j] \leftarrow \frac{R_{split}[i,j]}{Q_{split}[i,j]}$
20:     **end for**
21: **end for**
22: merger $\leftarrow \text{merger}(\text{normalize}(P_{merger}))$
23: split $\leftarrow \text{split}(\text{normalize}(P_{split}))$
24: novelty $\leftarrow \text{novelty}(Q)$
25:
26: **return** merger, split, novelty

seeds and their associated Wikipedia pages, it is possible to extract the whole textual content and build artificial documents for the chosen concepts[3].

In the following exercise, we selected 8 concept seeds, all related to the field of Economics, in order to understand how well our approach works on a toy model reflecting contents which are consistent with the real data we used in Section 4).

As in most natural language processing systems, we applied some pre-processing phase, which includes the removal of stopwords as well as functional linguistic items such as determiners, punctuations, etc[4].

Once the sets of words are built, we generated a document for each seed concept by randomly selecting the words[5] with uniform probability. We maintained word repetitions to allow us to sampling words with their real frequency and generate documents closed to real cases. The documents

---

[3]We used the library *Wikipedia* available at https://github.com/goldsmith/Wikipedia, which acts as a *w*rapper of the MediaWiki API (https://www.mediawiki.org/wiki/)

[4]We used the library *Spacy* (https://spacy.io/), filtering out the words having the following Part-of-Speech tags: DET (article), NUM (number) and PUNCT (punctuation).

[5]The number of words of each document has been chosen randomly.

generated are used to train different LDA models with different seeds concepts.

Finally, we compared the topics of different LDA models by means of the proposed measures to see whether they capture the dynamics of the topic changes. We refer the reader to Appendix A for details about the algorithms.

3.2   Controlled Experiments

To evaluate the algorithm we create 8 different controlled experiments which are designed to capture the 4 ideal cases of knowledge evolution. Specifically, we conducted twice 4 experiments to test the functioning of the method in 4 different situations by changing (or not) the number of topics and by replacing (or not) the concept seed. In the first 4 runs we kept the scenario as simple as possible and we slightly increased the complexity of the exercise in the second 4 runs.

In the former, the number of topics at time t are fixed to 2 for the first experiment and 4 for the second one; the number of topics at time $t+1$ is determined by the experiment (see Table 4 for details). In details, we set each experiment as follows:

stability the number of topics and seed concepts are kept the same. The variation is only stochastic.

birth/death the number of topics does not change, but we replace the concept seeds to force death of the previous topics and birth of new ones.

merging the seed concepts do not change, but we reduce the number of topics to force a situation of merging. For instance, if we cluster the same concept seeds in 2 and 1 topics, we necessarily observe only merging and no splitting.

splitting the seed concepts do not change, but we increase the number of topics to force a situation of splitting.

Table 4 summarizes the design of the experiments and depicts avarage values of 100 runs of the *Algorithm 2*. Concerning with the first 4 simple designs, experiments are conceived to force the results and create only splitting and only merging. For the splitting the number of topics increases from one to two and we should not observe merging since at $t-1$ there is also one topic. Analogously, in the case of merging the number of topic shrinks to one in $t+1$. The remaining two experiments compare stability with births and deaths, which lead to a higher degree of novelty. Our indexes vary as expected: in *splitting* and *merging* the MI and SI respectively go to zero. If we compare *stability* with *births and deaths* the NI is much higher in the former case. Table 4 shows four different experiments with higher number of topics. It is relevant to notice that even with a few topics, it is impossible to get a clear-cut outcome since the recombination of knowledge may be unexpected and typically reproduces at the same time merging, splitting, stability for some topics, and birth and death for others. However, these baseline examples clearly points at the aggregate behaviour of topics within a discipline.

Table 4: The table shows the experimental results conducted over the four cases.

Type of experiment	Concept seed	Replacement	Topics at time t	Topics at time t+1	index	value
stability	labour economics innovation economics	no	2	2	MI SI NI	0.189 0.192 0.43
splitting	labour economics innovation economics	no	1	2	MI SI NI	0.0 0.328 0.155
merging	labour economics innovation economics	no	2	1	MI SI NI	0.398 0.0 0.296
birth/death	labour economics innovation economics	cultural economics environmental economics	2	2	MI SI NI	0.363 0.385 0.768
stability	labour economics innovation economics cultural economics environmental ecs	no	4	4	MI SI NI	0.541 0.545 0.555
splitting	labour economics innovation economics cultural economics environmental ecs	no	4	8	MI SI NI	0.574 0.747 0.561
merging	labour economics innovation economics cultural economics environmental economics	no	4	2	MI SI NI	0.587 0.297 0.472
birth/death	labour ecs innovation economics cultural economics environmental economics	industrial economics transport economics economic history health economics	4	4	MI SI NI	0.640 0.662 0.777

4 The Evolution of Knowledge in Economics

The dataset is a collection of documents which appear in the JSTOR database (`www.jstor.org`) and were published from 1845 to 2013 in more than 190 journals concerning with economic sciences (also defined as *economics*). They are more than 460,000 documents, classified as research articles (about 250,000), book reviews (135,000), miscellaneous (73,000), news (4,000) and editorials (500). For each document, in addition to bibliographic information (title, publication date, authors, journal title, etc.), the dataset provides full content in form of a bag of words, i.e. the set of words used in the documents associated with their frequencies.

The following analysis only considers the research articles in order to remove the possible noise caused by using different types of documents, which can be written in different languages. The distribution of research articles over the time considered is very skewed (see Figure 5). Although the first documents date back to 1845, until the end of the XIX century the corpus of articles accounts only for 2930 items. The increase is almost linear till the beginning of the 1960s, when the number of documents more than doubled in a few years and rose to over 5000 items published every year during the 1990s and 2000s. From 2011 to 2013 we count 8220 items published.

The LDA has been applied to research papers published between 1890 and 2013: decades before 1890 were dropped because of the extremely low number of documents. Thereby, the resulting dataset of articles consists of 755,838,336 words and 3,169,515 unique words. We experimented varying the hyper-parameters of the method, namely the number of topics and the dimension of time windows, in order to evaluate the robustness and sensitivity of our approach in the 123 years considered. We selected 25, 50 and 100 topics and time windows of 5, 10 and 20 years, keeping fixed one parameter and varying the other one. In details, we first show the values of *SI* and *MI* fixing the window dimension to 10 years and varying the number of topics. In the following figures, for example, 1900-1920 indicates the value of the indexes between 1900 and 1910 compared with the corresponding value between 1910 and 1920. Figures 6, 7 and 8 show the indexes for 25, 50 and 100 topics within a window of 10 years. Then, we fixed the number of topics to 25 and we varied the size of the time window. Figures 9 and 10 show the indexes for 25 topics and windows of 5 and 20 years.

These simple tests have demonstrated that the main trends of the indexes do not change substantially by varying the hyper-parameters, meaning that our method is robust to the number of topics and the size of the time windows.

To further prove invariance to number of topics and windows-size, we applied Greene metric Greene et al. [2014] on a subset of the research articles with a time windows of 10 years to capture all the possible changing in economic knowledge. Values of the metric reveal how much the topics generated capture the information presented in the dataset. Greene metric requires a range in input, which is formed by the minimum and maximum number of topics, and a step parameter, used by the metric to shift the number of topics considered at the current step starting from the minimum ones. For example, if the minimum number of topics is 10, the maximum is 50 and the step is 20, the Greene metrics will compute a score at 10, 30 and 50 topics. The plot of the metric in Figures 11 and 12 concerns with two windows and shows that increasing the number of topics we can increase stability too, but of course, it becomes very difficult to interpret the meaning of each topic.

As suggested by Mimno and Blei [2011] when topic modeling is employed to explore the content of a dataset -as in this paper - rather than to predict there is not a definitive test to support the choice of the optimal number of topics. We solved this trade-off between stability and meaningfulness by

Figure 5: Distribution of documents in the corpus per year of publication
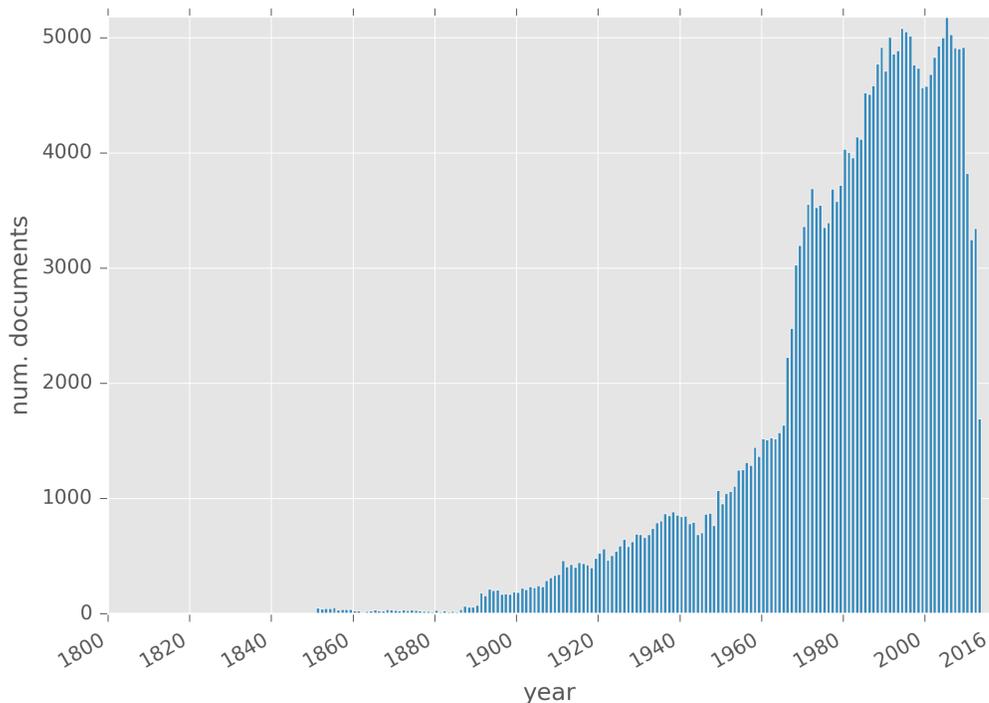
manually controlling for the topics generated by the model with 25 topics within time window of 10 years. When we found that a few topics could be split up again because they were too general, we set an optimal and analytically useful number of topics to 27. Therefore, the following analysis is based on 27 topics within time windows of 10 years, which perform the maximum stability of the indexes varying the number of topics.

Figure 13 shows the values of MI and SI respectively for each time window, as defined in Section 2.2. In the corpus we analyzed, both indexes show a general trend of decreasing values over time, which becomes particularly severe starting from 1960s. Merging and splitting increase only between the 1940s and the 1950s while dropping dramatically in the second half of the XX century. The transformation of topics seems to find new urge only around the end of the century, when merging is increasing againg and splitting is stable. As for the NI, we mentioned that the index tends to one when new topics emerge without matching with topics at $t-1$. On the average the value is higher than 0.9 all over the 123 years considered, so we tracked both micro-variation and general trend. In Figure 14 NI does not show relevant variations until 1990s, with some local maximum in the first decade of the past century and a local minimum around the half of it. In the last decade
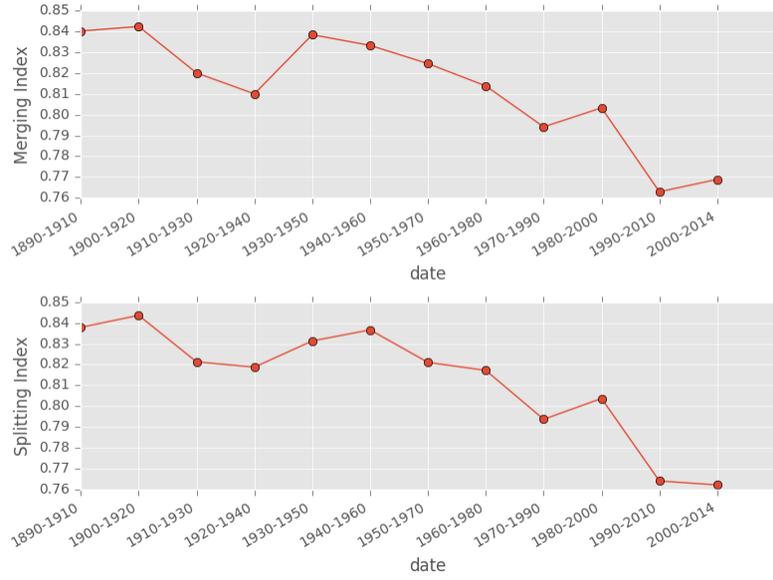
15

Figure 6: MI and SI - 25 topics 10 years window size
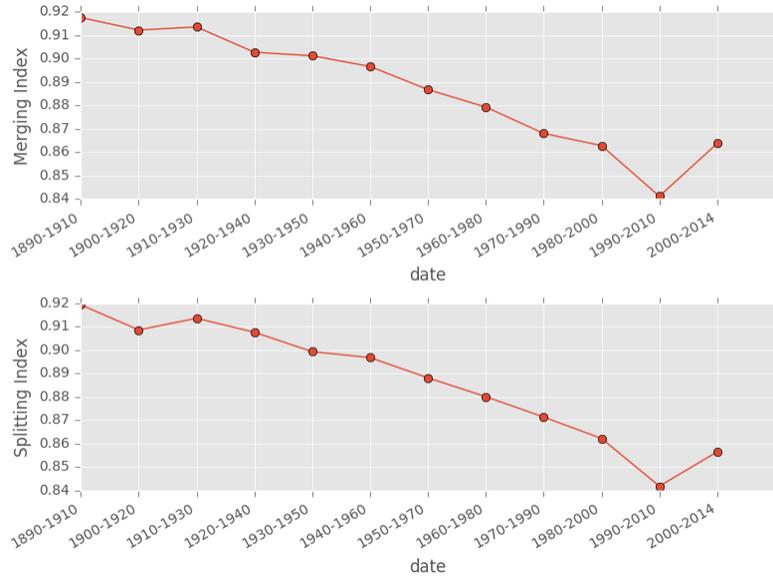

Figure 7: MI and SI - 50 topics and 10 years window size

of the century it grows sharply, revealing a higher rate of brand new topics or at least of topics defined by new words.
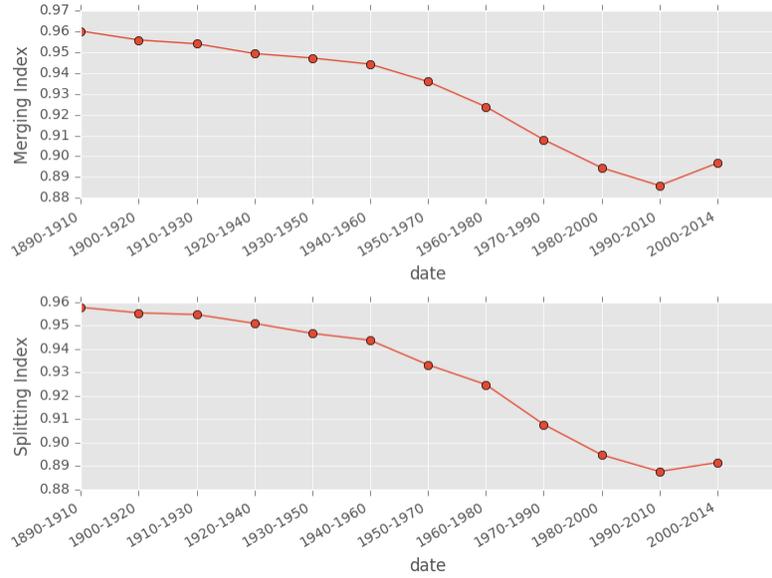
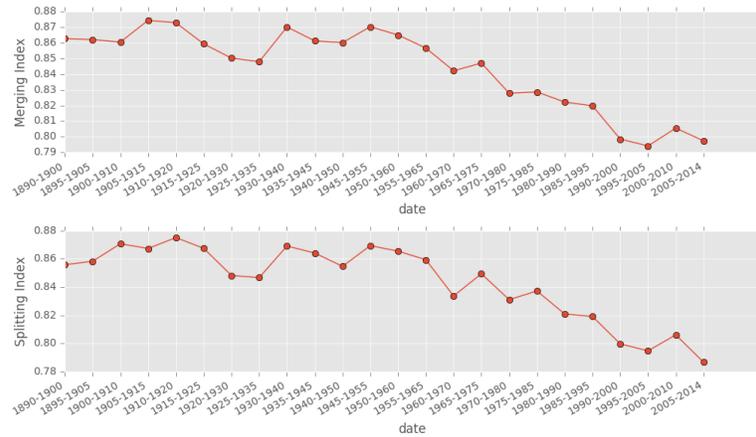Figure 8: MI and SI - 100 topics and 10 years window size

Figure 9: MI and SI - 25 topics and 5 years window size

Such a methodological approach has the advantage of tracking the evolution of each single stream of economic theory by looking simultaneously at all the others. On the whole, the analysis of such a big corpus of documents suggests that merging and splitting cannot be considered as opposite phenomena, but a complementary measure of recombination of topics. In particular, trends in the field of economics suggest a steady decrease of both splitting and merging only temporally balanced by a weak growth before and after the WWII. From a historical perspective this is absolutely consistent with the need of theoretical elaboration in economics following the great Depression in
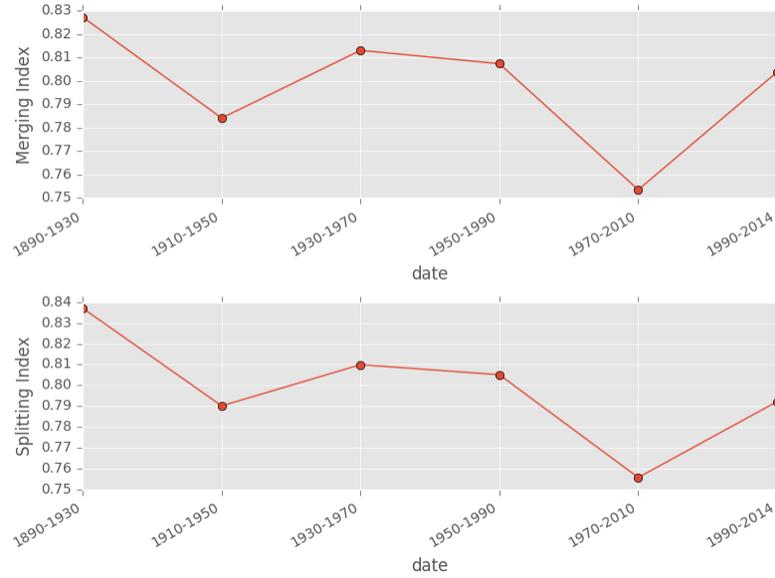
17

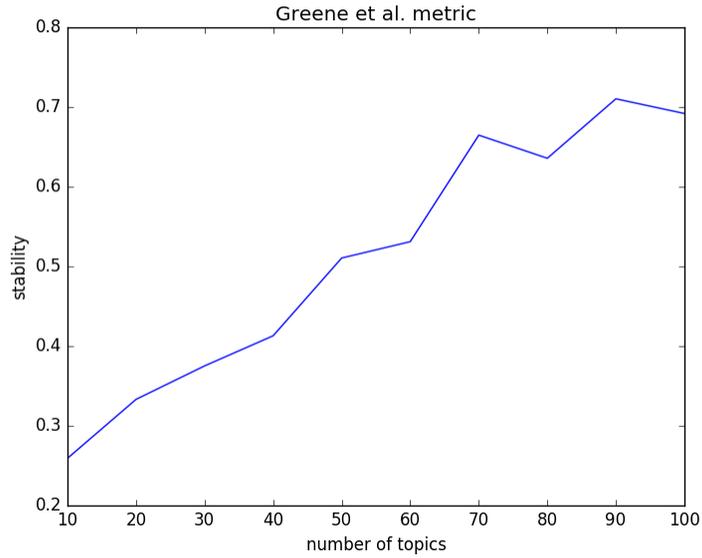Figure 10: MI and SI - 25 topics and 20 years window size

Figure 11: Greene et al.'s stability values for the time window 1910-1920.

1929 and the dramatic economic changed imposed by the post war reconstruction. During the 1960s and in combination with the boom of academic publications, many topics are spreads over
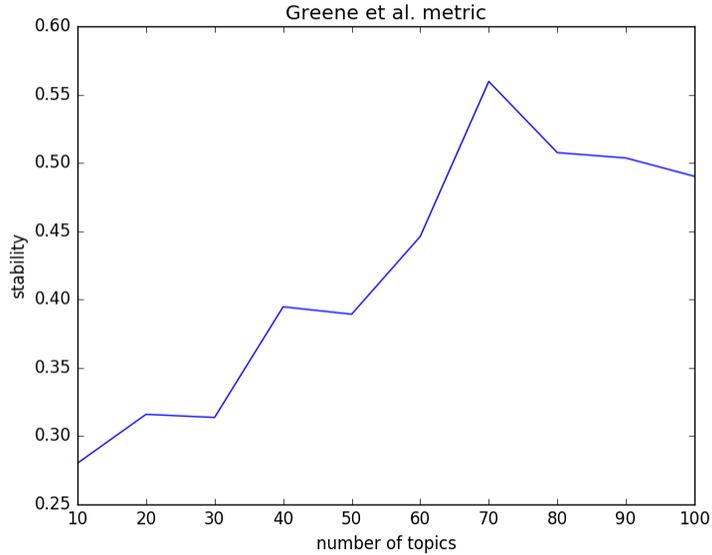
18

Figure 12: Greene et al.'s stability values for the time window 1940-1950.

a relevant number of documents and journals, although they seem to elaborate on relative stable basis of autonomous topics. Only by the end of the century we have witnessed the development of new-brand topics. The birth of new topics strengthens the hypothesis of self-standing topics shaped by their own specialised language and a lesser exchange of knowledge across the economic discipline. In other words, the terrific expansion of the academic production seems to come with a fragmentation and dispersion in multiple niches of knowledge Cedrini and Fontana [2017] which elaborate on a new language, but not necessarily producing new paradigms.

5   Conclusion

In this paper we proposed a method to measure the evolution of knowledge in a scientific field extracting topics in a corpus of documents. Topic modeling techniques are becoming increasingly refined in treating large and complex corpora of documents, but they may lack of a theoretical reflection of the underlying empirical phenomenon. Taking a dynamic perspective we recognise five paradigmatic cases of knowledge evolution. We then surmise that modeling the proximity between topics of different time windows as a proximity network might be a useful tool to measure their knowledge dynamics. Indeed, this network approach allows us to develop 3 indexes, which grasp i) the stability of topics over time measuring their rate of death and birth (*Novelty Index - NI*), and ii) the degree of recombination of topics (*Merging Index - MI* and *Splitting Index - SI*). For very simple cases, we are also able to analytically derive those conditions, which link the proximity network and the value of each index. Testing the algorithm over a set of simulated documents, we showed its robustness for each the indexed developed. Finally, we applied our approach to a real and large corpus of academic publications in economics to illustrate how the combined use of *MI*,

19

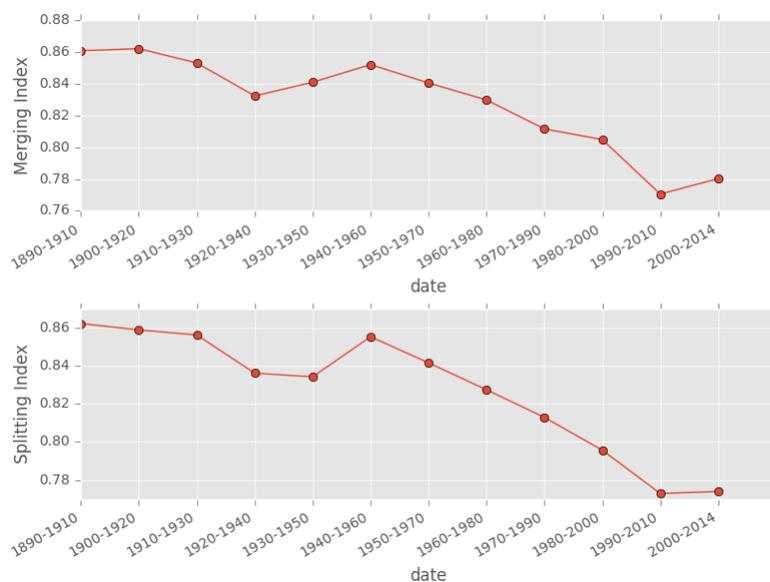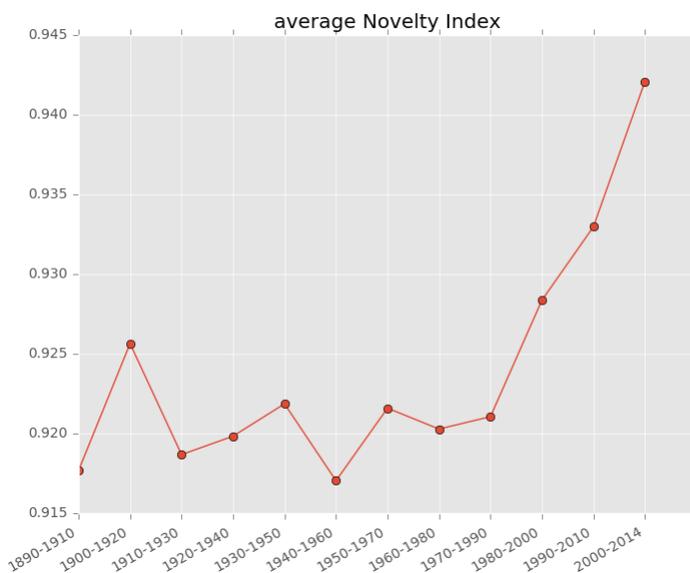Figure 13: MI and SI - 27 topics 10 years

Figure 14: NI - 27 topics 10 years

SI and NI is effective to understand dynamics and trends in economic knowledge and thought. We believe, this is a first step towards the development of a closer connection between algorithms
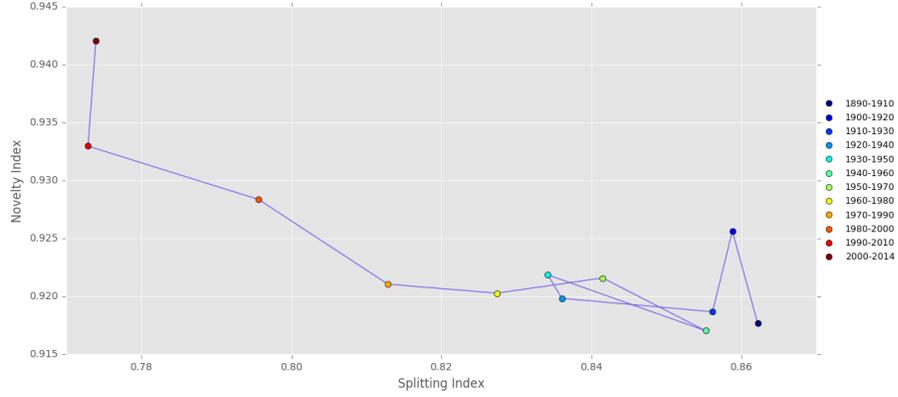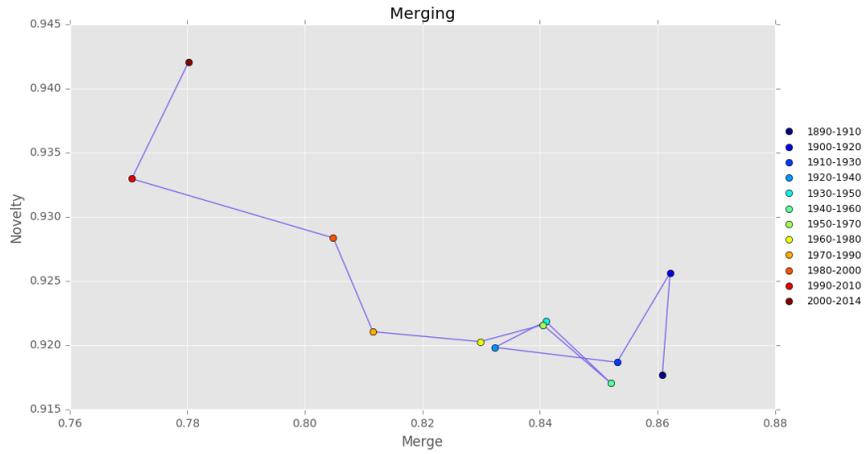
20

Figure 15: Combined graph of SI and NI

Figure 16: Combined graph of MI and NI

for dynamic topic modeling and the empirical phenomenon they are supposed to describe.

A   Artificial Data Creation: Algorithms

In *Algorithm 2*, the function *getNum(minNum,maxNum)* returns a number, randomly selected, between *minNum* e *maxNum*; the *getWord()* function returns a word, randomly chosen on the selected set; the function *computeTopicSimilarity()* calculates the cosine similarity between the input topics; the function *zeros()* returns an array containing all zeros. Finally, the function *getWordList(concept)* generates a set of words. The words are taken from the Wikipedia page that points to the chosen concept.

In rows [1-6], the function *getWordList* collects, for each concept seed, a set of words. In details, *getWordList*, as shown in *Algorithm 3*, extracts all words contained both in the Wikipedia page related to the concept in input through the python library *Wikipedia*[6]. Words are extracted using the library *Spacy*[7] and stored in *wordList*[8]. Then, the *wordList* of each *concept seed* is inserted into *wordConceptList*. In rows [7-16], *Algorithm 2* generates a document for each concept, sampling words (with uniform probability) from the *wordList* related to the *concept seed*. The number of words to sample is specified by *numWords*, which ranges from 1000 to 10000. Successively, in rows [18-20], the algorithm divides documents in two sets, a set containing the first *numDocument* documents and a set containing the remain documents, and applies LDA. The LDA can be applied over the two documents sets or only over a single documents set according to the *replaceDoc* flag. If *replaceDoc* is set to *True*, the first documents set is replaced with the second one (it is set to *False* by default).

Algorithm 4 shows how words are processed. We filtered stopwords and words having *Part-Of-Speech* tags *Det* (Determiner), *X* (foreign word), *NUM* (Numeral), *Punct* (Punctuation), *SPACE* and *EOL* (end of line symbols). We also filtered words that does not match the python regular expression $\backslash w+$. Furthermore, all unfiltered words are brought back to their morphological root.

---

[6] https://github.com/goldsmith/Wikipedia

[7] https://spacy.io/

[8] There exists a *wordList* for each *conceptSeed* in input.

Algorithm 2 ToyEvaluation(seedConcepts, numDocument, $numTopic_t$, $numTopic_{t+1}$, replace-Doc)

1: wordsConceptList = {}
2: // create a words list for each concept seed
3: **for** concept in seedConcepts **do**
4:     wordsList ← getWordList(concept)
5:     wordsConceptList.append(wordsList)
6: **end for**
7: documents = {}
8: **for** i ← 1..len(seedConcepts) **do**
9:     numWords ← getNum(1000, 10000)
10:     document = {}
11:     **for** j ← 1..numWords **do**
12:         word ← wordsConceptList[i].getWord()
13:         document.append(word)
14:     **end for**
15:     documents.append(document)
16: **end for**
17: // get topic
18: documentSet ← documents[1:numDocument]
19: $topic_t$ ← LDA(documentSet, $numTopic_t$)
20: M_t ← computeTopicSimilarity($topic_t$, $topic_t$)
21: **if** replaceDoc ≠ False **then**
22:     documentSet ← documents[numDocument:len(seedConcepts)]
23: **end if**
24: $topic_{t+1}$ ← LDA(documentSet, $numTopic_{t+1}$)
25: M_{t+1} ← computeTopicSimilarity($topic_{t+1}$, $topic_{t+1}$)
26:
27: /* it then continues as computeSingleWindow algorithm */

Algorithm 3 getWordList(concept)

1: posTags ← {X, NUM, DET, PUNCT}
2: parser ← parser(lan=eng)
3: wordList ← {}
4: wordList ← getWordList(content, posTags)
5: **return** wordList

Algorithm 4 getWords(content, posTags)

1: words ← {}
2: wikiPage ← Wikipedia.getPage(concept)
3: **for** sentence in parser(wikiPage.content).sentences **do**
4:   **for** word in sentence.words **do**
5:     **if** ¬(word in stopwords) ∧ ¬(word.pos in posTags) ∧ match(word,\w+) **then**
6:       words.append(word.lemma)
7:     **end if**
8:   **end for**
9: **end for**
10: **return** words

References

R. Alghamdi and K. Alfalqi. A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications*, 6(1):147–153, 2015.

D. M. Blei and J. D. Lafferty. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 113–120, New York, NY, USA, 2006. ACM. ISBN 1-59593-383-2. doi: 10.1145/1143844.1143859. URL http://doi.acm.org/10.1145/1143844.1143859.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3:993–1022., January 2003.

L. Bolelli, S. Ertekin, D. Zhou, and C. L. Giles. Finding topic trends in digital libraries. In *Proceedings of the 9th ACMIEEE-CS joint conference on Digital libraries*, pages 69–72. ACM, 2009.

M. Callon, J.-P. Courtial, W. A. Turner, and S. Bauin. From translations to problematic networks: An introduction to co-word analysis. *Social science information*, 22(2):191–235, 1983.

M. Cedrini and M. Fontana. Just another niche in the wall? how specialization is changing the face of mainstream economics. *Cambridge Journal of Economics*, forthcoming, 2017.

P. DiMaggio, M. Nag, and D. Blei. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics*, 41(6):570–606, 2013.

D. Greene, D. O'Callaghan, and P. Cunningham. *How Many Topics? Stability Analysis for Topic Models*, pages 498–513. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. ISBN 978-3-662-44848-9. doi: 10.1007/978-3-662-44848-9_32. URL http://dx.doi.org/10.1007/978-3-662-44848-9_32.

Q. He, B. Chen, J. Pei, B. Qiu, P. Mitra, and L. Giles. Detecting topic evolution in scientific literature: how can citations help? *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 957–966, November 2009.

M. I. Jordan. *Learning in graphical models*, volume 89. Springer Science & Business Media, 1998.

T. S. Kuhn. *The structure of scientific revolutions, International Encyclopedia of Unified Science, vol. 2, no. 2.* Chicago: The University of Chicago Press, 1970.

L. Laudan. *Progress and its problems: Towards a theory of scientific growth.* Univ of California Press, 1978.

S. L. Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

L. Leydesdorff and A. Nerghes. Co-word maps and topic modeling: A comparison from a user's perspective. *arXiv preprint arXiv:1511.03020*, 2015.

L. Leydesdorff and K. Welbers. The semantic mapping of words and co-words in contexts. *Journal of Informetrics*, 5(3):469–475, 2011.

D. Mimno and D. Blei. Bayesian checking for topic models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 227–237, 2011.

M. F. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

A. Suominen and H. Toivanen. Map of science with topic modeling: Comparison of unsupervised learning and human?assigned subject classification. *Journal of the Association for Information Science and Technology*, October 2015.

E. Vlieger and L. Leydesdorff. Content analysis and the measurement of meaning: The visualization of frames in collections of messages. *Public Journal of Semiotics*, 3(1):28–50, 2011.