

Systematic Litterature Review

Christophe Benavent et Olivier Caron

Université Dauphine-PSL - DRM - Acss

February 10, 2023

Sommaire I

- 1 Introduction
- 2 Data sets acquisition
- 3 Network analysis
- 4 Topic Models and embeddings
- 5 Conclusion
- 6 References

Section 1

Introduction

Introduction

Systematic reviews are different from traditional literature reviews because they aim to identify all studies that address a specific question. Typically : “How much high are advertising elasticities at the level of the brand.

- so not only a quantitative dimension cause large amount of references.
- Need Text mining methods because we deal mainly with text, focusing on values would be a meta-analyses.
- Toward reproducibility criteria.

In this presentation a pragmatic approach through three questions

- How to constitute the data set ?
- How analyse Authorship and communities ?
- How to Analyse contents ?

Narratives versus systematic ?

Type	Advantage(s)	Disadvantage	Application(s)	Guidelines
Systematic Review	<ol style="list-style-type: none"> 1. Minimized bias 2. A-priori protocol 3. Defined search and evaluation methods 4. Reproducible 5. High validity of review conclusions 	<ol style="list-style-type: none"> 1. Must adhere to established guidelines 2. Valid literature base required 3. Robust (enough) literature to review 4. Variation in study methods within reviewed literature may affect results 	<ol style="list-style-type: none"> 1. Identify relevant evidence 2. Assess quality of evidence 3. Non-biased synthesis of literature 4. Interpret evidence in an impartial manner 5. Applicable for establishing standards and health policy 	PRISMA Guidelines ²
Meta-Analyses - Quantitative	<ol style="list-style-type: none"> 1. Same as systematic review 2. Determine a single estimate of the effect of treatment or management of an illness or event 	<ol style="list-style-type: none"> 1. Data in literature must be homogeneous and available for pooled analysis 2. Reliability of literature designs may affect results 	<ol style="list-style-type: none"> 1. Same as systematic review 2. Determine best practice for defined topic or event. 3. Narrow variations in known data sets. 	PRISMA Guidelines ²
Meta-Analyses - Qualitative	<ol style="list-style-type: none"> 1. Same as systematic review 2. Determine major themes or experiences for an event or issue 	<ol style="list-style-type: none"> 1. Variable sampling errors in original literature leads to bias 2. Variation in qualitative tools used for original research 	<ol style="list-style-type: none"> 1. Same as systematic review 2. Define primary themes and priorities 3. Refine future research objectives 	PRISMA Guidelines ²
Cochrane Review	<ol style="list-style-type: none"> 1. Form of systematic review method 2. Well defined methodology 3. Indexed in the Cochrane Library (open source) 	<ol style="list-style-type: none"> 1. Same as for Systematic Reviews 	<ol style="list-style-type: none"> 1. Same as systematic review 2. Determine support for specific treatment 3. Determine if evidence exists for defined concept 	Cochrane Manual ⁶
Scoping Review	<ol style="list-style-type: none"> 1. Use of fluid literature search strategy 2. Broader review topics 3. May include literature of varied methodologies 	<ol style="list-style-type: none"> 1. Risk of bias due to lack of defined evaluation methods 2. Non-specific objectives 3. Heterogeneity in literature included 	<ol style="list-style-type: none"> 1. Map available literature in a review field or area 2. Literature gap analysis 3. Clarification of concept or theory 	PRISMA ScR ⁷
Narrative Review	<ol style="list-style-type: none"> 1. Researcher determines literature to include 2. Less time intensive 3. May include literature of varied methodologies 4. Interpretive objectives (not structured analysis) 	<ol style="list-style-type: none"> 1. Risk of multiple forms of bias and error 2. Unstructured, not reproducible 3. May not include all appropriate literature 4. Lacks systematic synthesis of literature 	<ol style="list-style-type: none"> 1. Identify theory and frames of thought on a topic 2. Summarize a particular study topic 3. Justify a research topic 	
Critical Review	Same as Narrative Review	Same as Narrative Review	<ol style="list-style-type: none"> 1. Develop perspectives on a topic 	
Conceptual Review	Same as Narrative Review	Same as Narrative Review	<ol style="list-style-type: none"> 1. Evaluate general consensus on a topic 2. Show gaps of knowledge in literature 	
State-of-the Art Review	Same as Narrative Review	Same as Narrative Review	<ol style="list-style-type: none"> 1. Describe current beliefs on a topic 	

Stratton © 2019 Psychosocial and Disaster Medicine

Figure 1: Narratives, systematic and others LR Stratton (2019)

the prisma framework

A systematic approach to set the article corpus Moher et al. (2009)

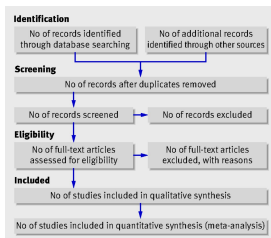


Figure 2: Figure 2 : Prisma Process

Section/Topic	Item #	Checklist Item	Reported on Page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome-level assessment (see Item 12).	
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot.	
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression (see Item 16)).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., health care providers, users, and policy makers).	
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias).	
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	

White and gray

Not only reviewed papers, but also working paper preprints etc.

Collecting with IA

- Elicit
- Litmaps
- Connected papers

r environnement

- r + Rstudio + Quarto/beamer to produce this presentation and doing computations.
- You can clone it at [github/benaventc](https://github.com/benaventc).
- Main Packages

```
library(tidyverse)
library(Rtsne)
library(ggrepel)
library(ggwordcloud)
library(quanteda)
library(quanteda.textstats)
library(quanteda.textmodels)
library(word2vec)
library(ape)
library(cowplot)

theme_set(theme_minimal()+theme(plot.title = element_text(size=12)))
```


Section 2

Data sets acquisition

Data sets acquisition

- Through database, with format (Bib, RIS, json), and API.

Some Sources

- Google Scholar : harvesting every things
- Crossref : open source
- Scopus : elsevier papers
- Ebsco : business source complete
- Jstor
- NBER
- Arxiv and other Psyxiv ou socioxiv, don't forget HAL,

...

A selection Process

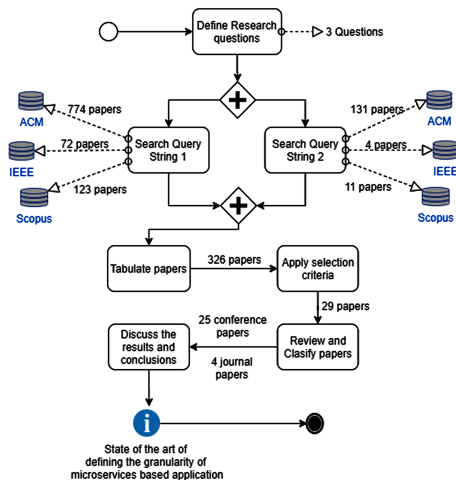
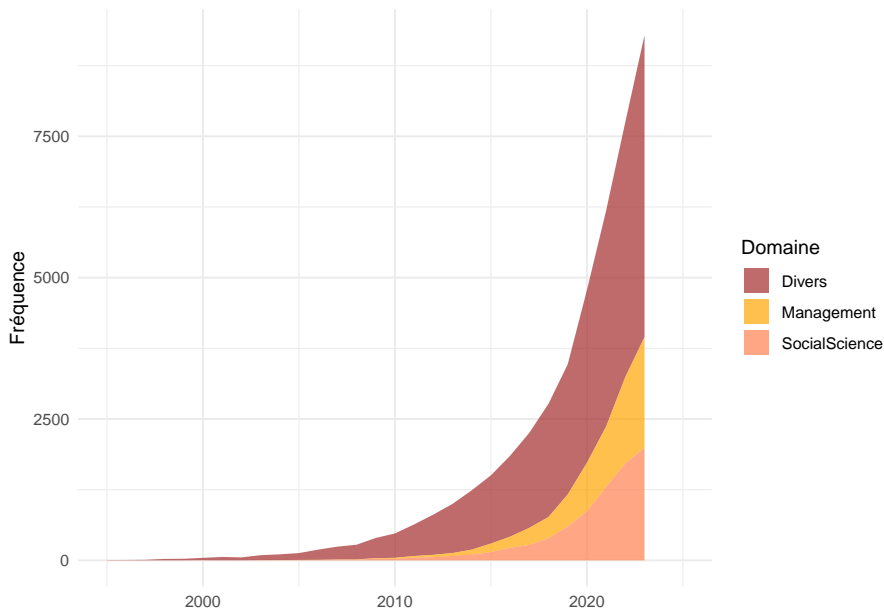


Figure 4: Figure 4 : An SLR process source : <https://doi.org/10.7717/peerj-cs.695/fig-1>

A short exemple with corpus

- **Scopus** : TITLE-ABS-KEY (“Systematic literature review”) -> 37,190 documents
- First health then computing science and a growing concern for social sciences.



A small Case study : NLP and marketing

A first case : “NLP in Marketing - state of art and evolutions” - Abstract, title, Keywords
=(“NLP” | “natural language processing” | “Text-Mining”| “text Analysis”) & Journal
=(“Marketing” | “Consumer”)

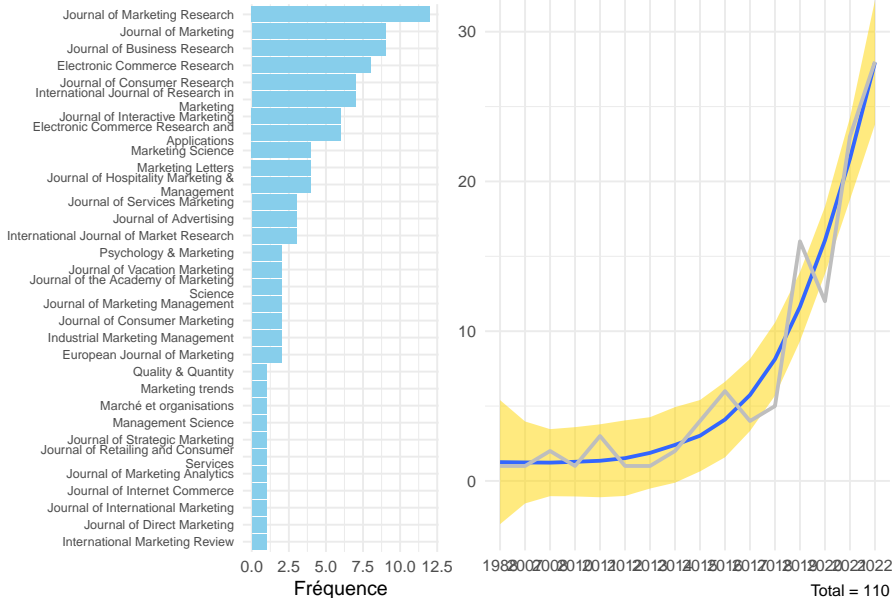
- The references are reported manually, through DOI and maintain in the Zotero collection, then export as datafile.
- Keywords are corrected and completed manually.
- Outcome : after cleaning -> 104 papers.

```
# A tibble: 5 x 9
```

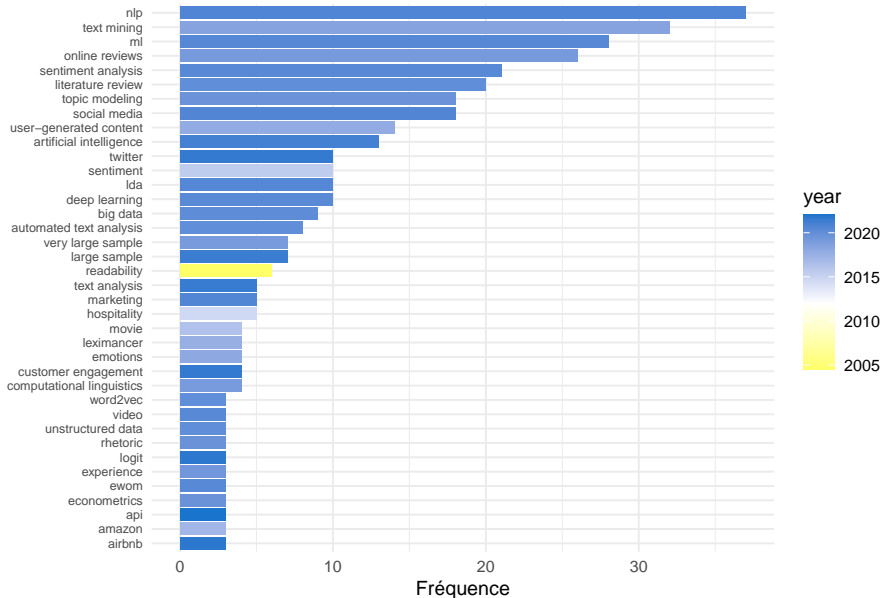
	id	year	auteurs	title	review	keywo~1	text	methods	fields
	<dbl>	<dbl>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	1	1988	Beard, John D.; William~	Incr~	Journ~	direct~	This~	Readab~	direc~
2	2	2007	Eliashberg, Jehoshua; H~	From~	Manag~	entert~	Movi~	predic~	movie
3	3	2008	Pekar, Viktor; Shiyao Ou	Disc~	Journ~	blogs,~	Auto~	opinio~	hospi~
4	4	2008	Sawyer, Alan G.; Laran,~	The ~	Journ~	readab~	This~	Readab~	publi~
5	5	2010	Nielek, Radoslaw; Wawer~	Spir~	Elect~	Auctio~	An a~	Sentim~	Aucti~

```
# ... with abbreviated variable name 1: keywords
```

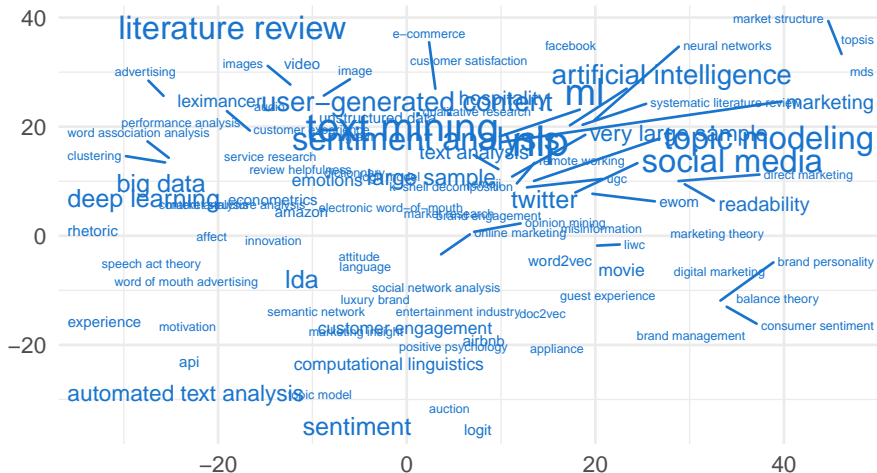
Marketing et NLP : Nombre d'articles par revue et par an



Mots clés les plus fréquents



Projection Tsne des mots clés



Using API

The best way to operate is to work through API :

- it prevents from errors
- it's precise
- it supports standard formats (bib, ris ...)

Main sources

- Crossref
- Scopus
- [Hal](<https://api.archives-ouvertes.fr/docs>)

Some libraries

- <https://www.bibliometrix.org/home/>
- <https://aurelien-goutsmedt.com/post/extracting-biblio-data-1/>
- <https://github.com/sbegueria/bibliometRics>
- searchlitr package r
- <https://rdrr.io/github/nfrerebeau/odyssey/f/README.md>
-

Focus on crossrefs

rcrossref : le plus important ? mais fermé à la liste bibliographique cite-by.

Section 3

Network analysis

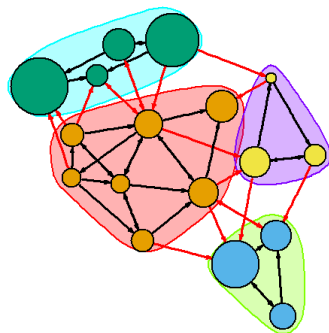


Figure 5: A network and communities

- *igraph* the perfect tool with r and (an excellent introduction)[<https://kateto.net/netscix2016.html>]
- data : $x < -w_i - > y$
 - ▶ co-occurrences and others distances.
- Analytical tools :
 - ▶ Layout (MDS, KR, ...)
 - ▶ Centrality measurements (HITS, ...)
 - ▶ Cliques and communities detection

PMP Authorship Analysis

- PMP case : 40 years of publications - around 1020 papers
- Tracking a regime change

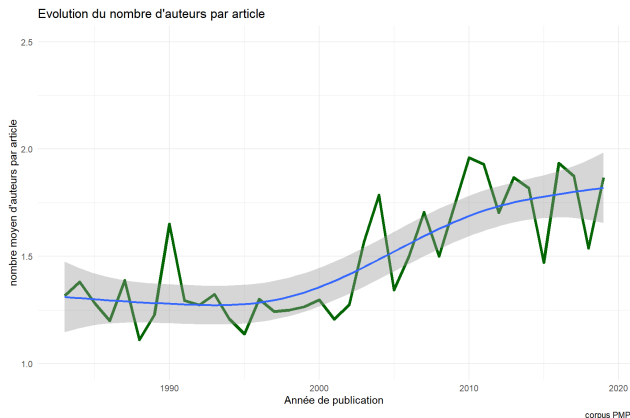


Figure 6: Figure 10 : a change in authorship

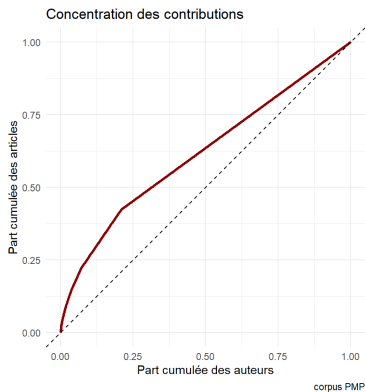
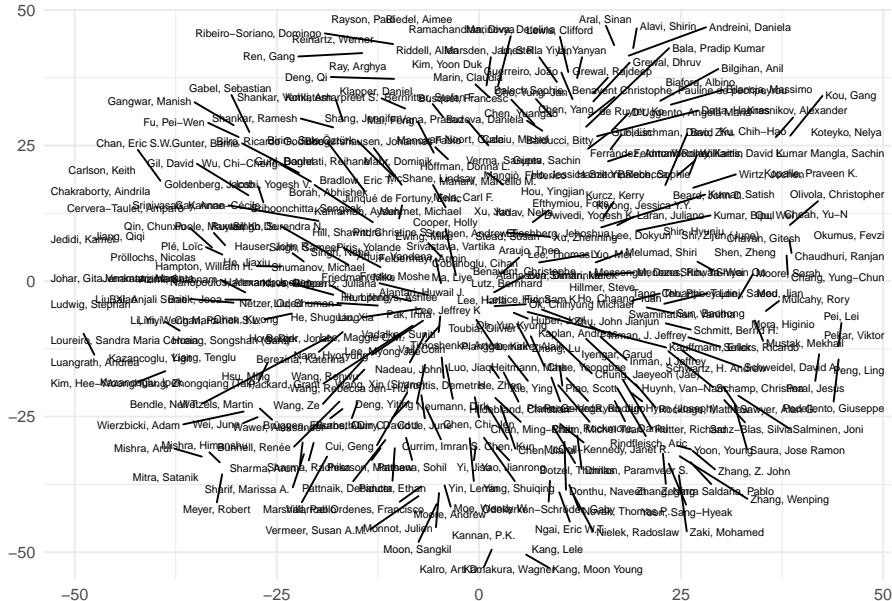


Figure 7: Author concentration

Projection Tsne des mots clés



L'exemple d'un réseau de citation

And

Un petit exemple par olivier

Section 4

Topic Models and embeddings

Topic Models and embeddings

- Topic model represent the first modern wave of text statistical modeling approach with LDA models Blei, Ng, and Jordan (2003).
- Embeddings a second wave with Mikolov et al. (2013)
- Transformers is all you need, now.

An application of a STM model

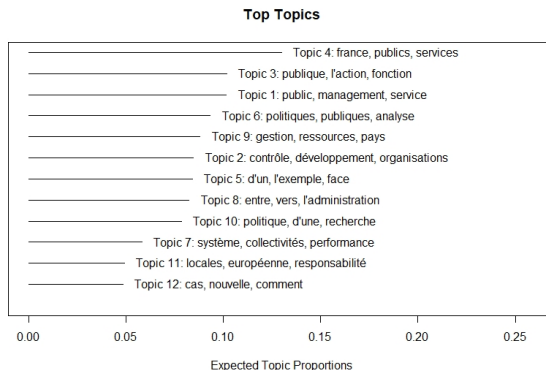


Figure 9: A STM topic model : keywords and proportion of the content /n (each document has a p probability to belong to the topic k)

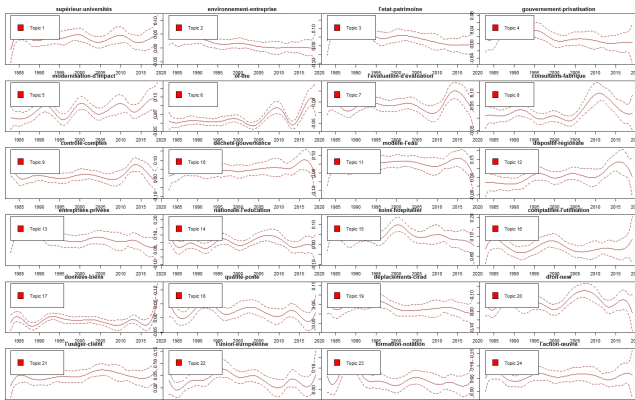


Figure 10: A STM topic model : time prevalence for each topic identified

An embeddings approach

On reprend le code marketing NLP

Section 5

Conclusion

Conclusion

- Another organisation of the literature review that requires a collective effort.
 - ▶ building common code
 - ▶ discussion arena
- Toward a systematic understanding of bibliometrics metadata
 - ▶ intrinsic : content and localization (Journal, date, institution)
 - first order information : pure meta data
 - second order information : outcomes, methods,
 - ▶ extrinsic : how papers related to the paper world
 - citation index,
 - co-references
- Be prepared for the disruption of deep NLP methods :
 - ▶ summarizing
 - ▶ concept extraction
 - ▶ outcome extraction
 - ▶ and more...

Section 6

References

References

There is also a set of full papers here

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation.” *J. Mach. Learn. Res.* 3 (March): 993–1022.

<http://dl.acm.org/citation.cfm?id=944919.944937>.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient Estimation of Word Representations in Vector Space.” arXiv. <http://arxiv.org/abs/1301.3781>.

Moher, D., A. Liberati, J. Tetzlaff, D. G Altman, and for the PRISMA Group. 2009.

“Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.” *BMJ* 339 (July): b2535–35. <https://doi.org/10.1136/bmj.b2535>.

Stratton, Samuel J. 2019. “Literature Reviews: Methods and Applications.” *Prehospital and Disaster Medicine* 34 (4): 347–49. <https://doi.org/10.1017/S1049023X19004588>.