Systematic Litterature Review

Christophe Benavent et Olivier Caron

Université Dauphine-PSL - DRM - Acss

February 16, 2023

Sommaire I

1. Introduction

2. Data sets acquisition

3. Network analysis

4. Topic Models and embeddings

5. Conclusion

Section 1

Introduction

Introduction

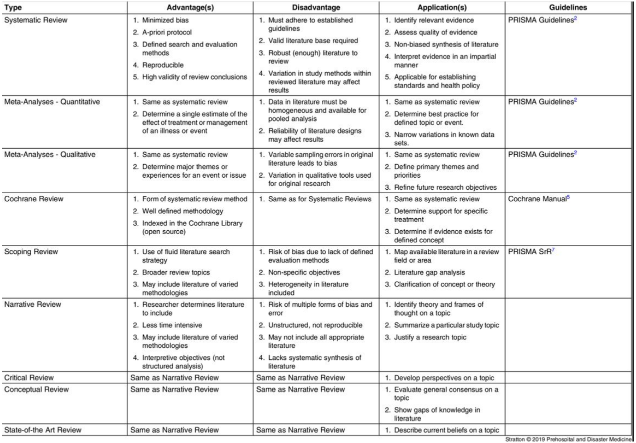
Systematic reviews are different from traditional literature reviews because they aim to identify all studies that address a specific question. Typically : “How much high are advertising elasticities at the level of the brand.

- 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 to analyse Authorship and communities ?
- How to Analyse contents ?

Narratives versus systematic ?



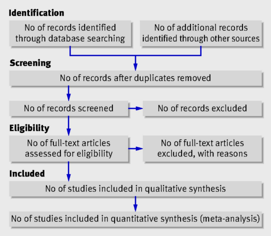
Type	Advantage(s)	Disadvantage	Application(s)	Guidelines
Systematic Review	1. Minimized bias 2. A-priori protocol 3. Defined search and evaluation methods 4. Reproducible 5. High validity of review conclusions	1. Must adhere to established guidelines 2. Valid literature base required 3. Robust (enough) literature to review 4. Variation in study methods within reviewed literature may affect results	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	1. Same as systematic review 2. Determine a single estimate of the effect of treatment or management of an illness or event	1. Data in literature must be homogeneous and available for pooled analysis 2. Reliability of literature designs may affect results	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	1. Same as systematic review 2. Determine major themes or experiences for an event or issue	1. Variable sampling errors in original literature leads to bias 2. Variation in qualitative tools used for original research	1. Same as systematic review 2. Define primary themes and priorities 3. Refine future research objectives	PRISMA Guidelines[2]
Cochrane Review	1. Form of systematic review method 2. Well defined methodology 3. Indexed in the Cochrane Library (open source)	1. Same as for Systematic Reviews	1. Same as systematic review 2. Determine support for specific treatment 3. Determine if evidence exists for defined concept	Cochrane Manual[5]
Scoping Review	1. Use of fluid literature search strategy 2. Broader review topics 3. May include literature of varied methodologies	1. Risk of bias due to lack of defined evaluation methods 2. Non-specific objectives 3. Heterogeneity in literature included	1. Map available literature in a review field or area 2. Literature gap analysis 3. Clarification of concept or theory	PRISMA SrR[7]
Narrative Review	1. Researcher determines literature to include 2. Less time intensive 3. May include literature of varied methodologies 4. Interpretive objectives (not structured analysis)	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	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	1. Develop perspectives on a topic	
Conceptual Review	Same as Narrative Review	Same as Narrative Review	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	1. Describe current beliefs on a topic	

Stratton © 2019 Prehospital and Disaster Medicine

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

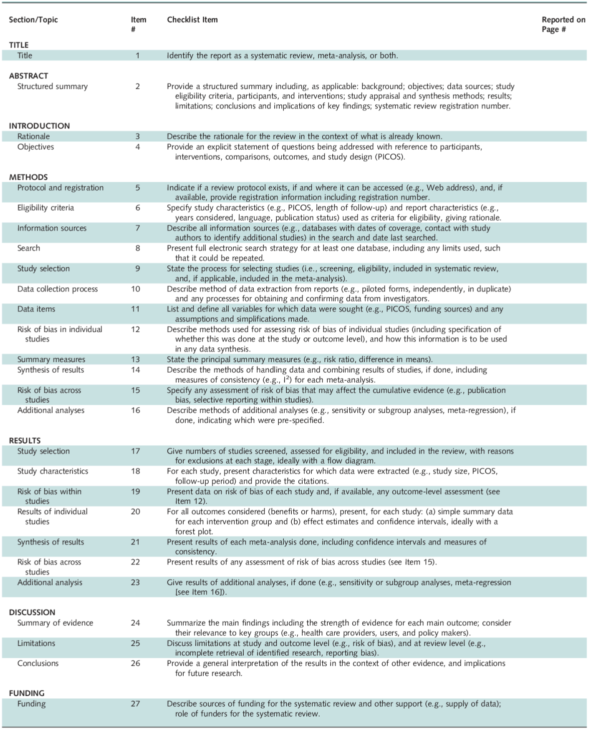
the prisma framework

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



Identification
No of records identified through database searching
No of additional records identified through other sources
Screening
No of records after duplicates removed
No of records screened
No of records excluded
Eligibility
No of full-text articles assessed for eligibility
No of full-text articles excluded, with reasons
Included
No of studies included in qualitative synthesis
No of studies included in quantitative synthesis (meta-analysis)

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.
- Main Packages

```r
library(tidyverse) # the tol Box
library(Rtsne)  # 2D magic
library(ggrepel)    #complement de ggplot
library(ggwordcloud)  #complement de ggplot
library(cowplot)

library(udpipe) # annotations
library(quanteda) # un bel ensemble de techniques
library(quanteda.textstats)
library(quanteda.textmodels)
library(quanteda.textplots)
library(fastcluster) #pour aller plus vite
library(ape) #phylo and clustering
library(word2vec) #for embeddings
library(rcrossref)
library(flextable)

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

Section 2

Data sets acquisition

Data sets acquisition

- Through citations 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

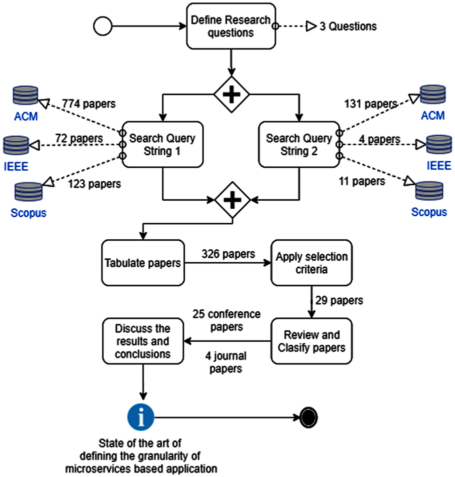
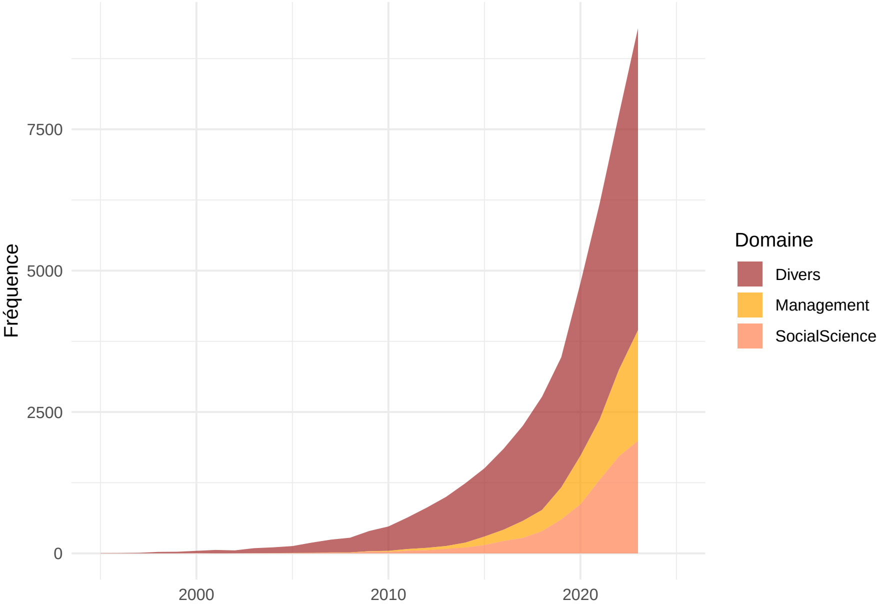


Figure 4: Figure 4 : An SLR process source : https://doi.org/10.7717/peerj-cs.695/fig-1

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.

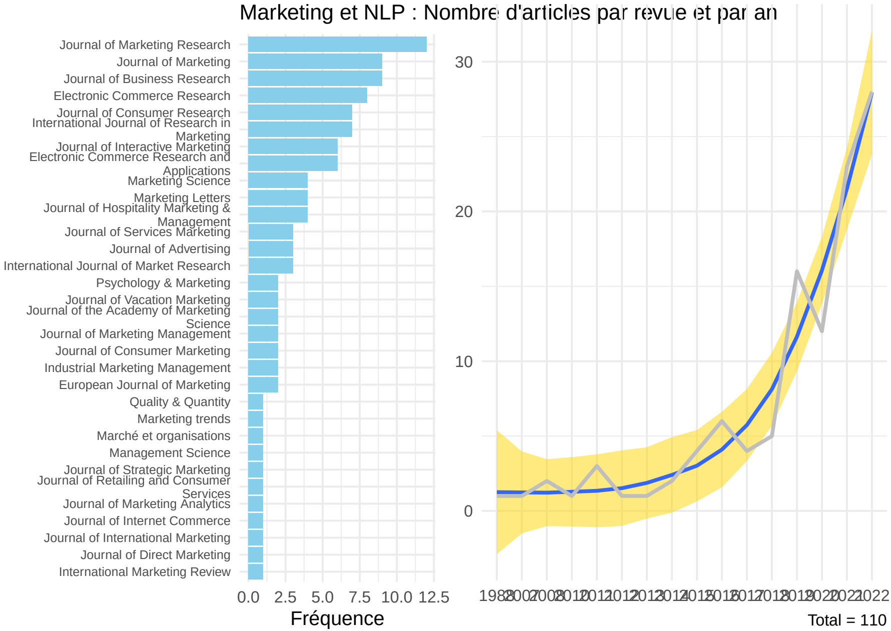


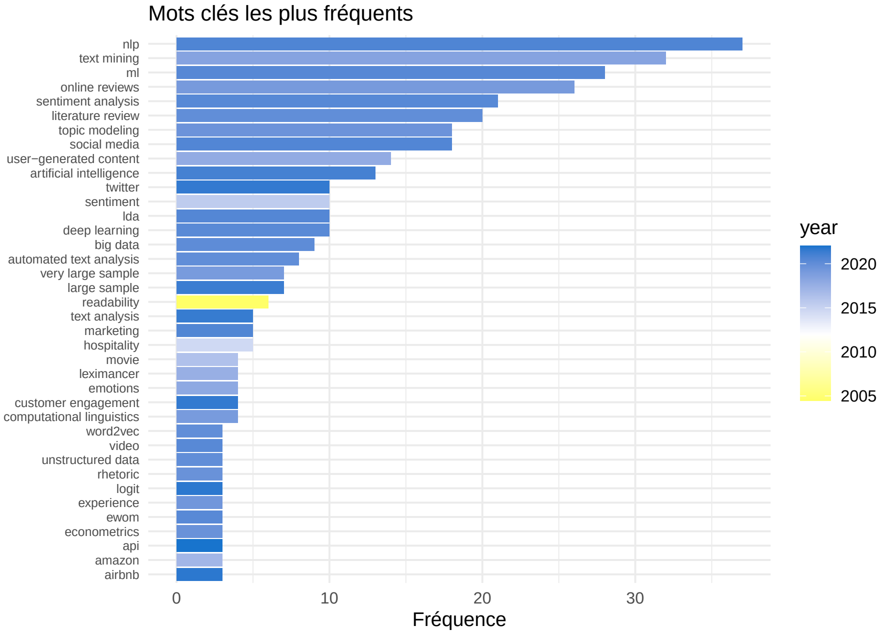
A small Case study : NLP and marketing

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”)

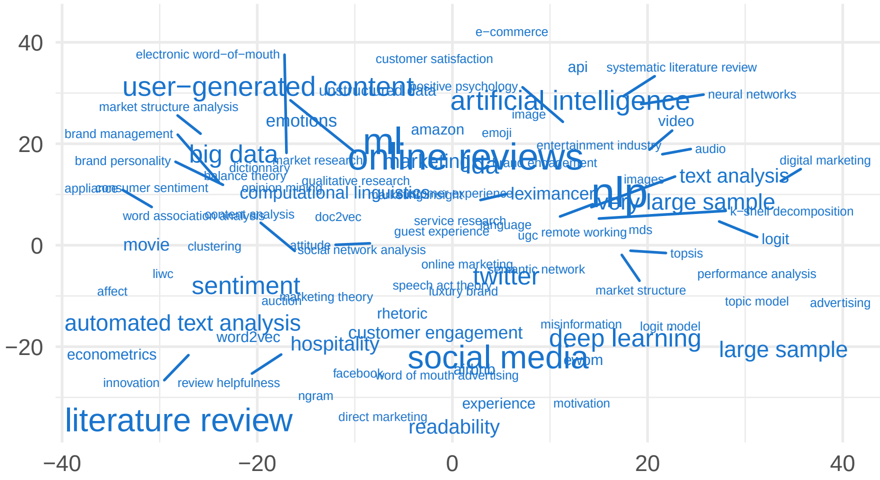
- The references are reported manually, through DOI and maintain in the Zotero collection, then export as datafile.
- Keywords are corrected and completed manually.
- Outcome : after cleaning -> 104 papers.

```
# 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; Shiyan 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
Journal of Marketing Research
Journal of Marketing
Journal of Business Research
Electronic Commerce Research
Journal of Consumer Research
International Journal of Research in Marketing
Journal of Interactive Marketing
Electronic Commerce Research and Applications
Marketing Science
Marketing Letters
Journal of Hospitality Marketing & Management
Journal of Services Marketing
Journal of Advertising
International Journal of Market Research
Psychology & Marketing
Journal of Vacation Marketing
Journal of the Academy of Marketing Science
Journal of Marketing Management
Journal of Consumer Marketing
Industrial Marketing Management
European Journal of Marketing
Quality & Quantity
Marketing trends
Marché et organisations
Management Science
Journal of Strategic Marketing
Journal of Retailing and Consumer Services
Journal of Marketing Analytics
Journal of Internet Commerce
Journal of International Marketing
Journal of Direct Marketing
International Marketing Review
0.0
2.5
5.0
7.5
10.0
12.5
Fréquence
30
20
10
0
1982
2007
2008
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
Total = 110


Mots clés les plus fréquents
nlp
text mining
ml
online reviews
sentiment analysis
literature review
topic modeling
social media
user−generated content
artificial intelligence
twitter
sentiment
lda
deep learning
big data
automated text analysis
very large sample
large sample
readability
text analysis
marketing
hospitality
movie
leximancer
emotions
customer engagement
computational linguistics
word2vec
video
unstructured data
rhetoric
logit
experience
ewom
econometrics
api
amazon
airbnb
0
10
20
30
Fréquence
year
2020
2015
2010
2005

Projection Tsne des mots clés



Using API

The best way to operate is to work through API :

- it it prevent from errors
- it's precise
- it support 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
- Fulltext package

Focus on crossrefs

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

Section 3

Network analysis

Network analysis

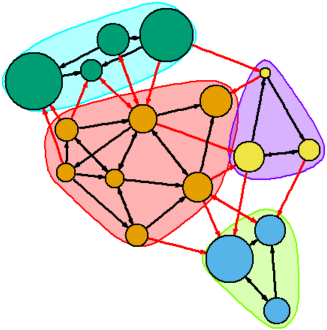


Figure 5: A network and communities

- *igraph* the perfect tool with r and (an excellent introduction)[https://kateto.net/netscix2016.html]
- data : $x < -w_i - > y$
  - co-occurences 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

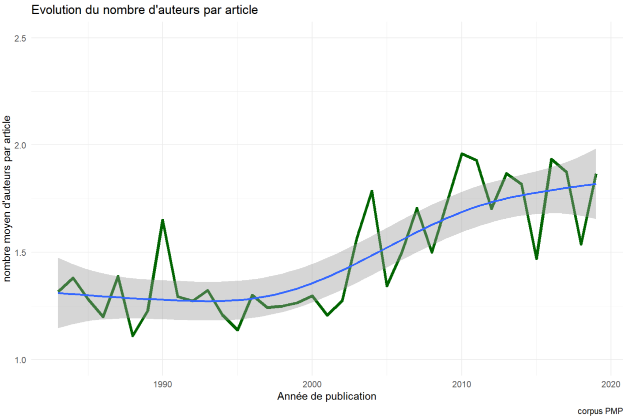


Figure 6: Figure 10 : a change in authorship

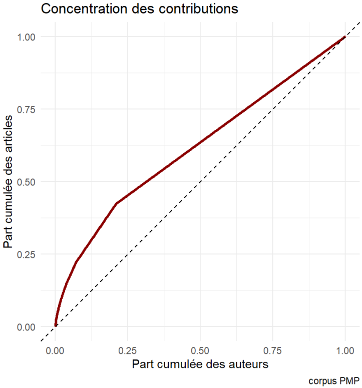
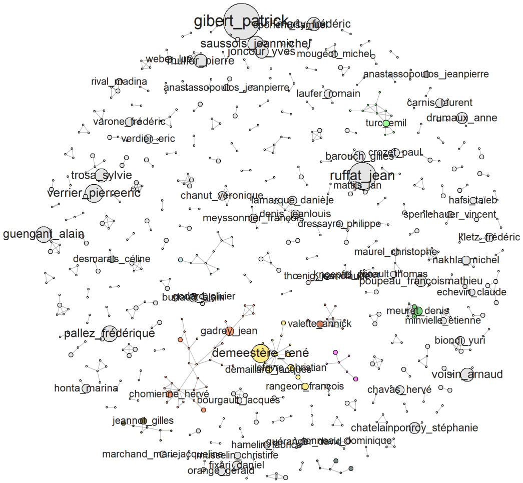


Figure 7: Author concentration

Les constellations des auteurs de PMP

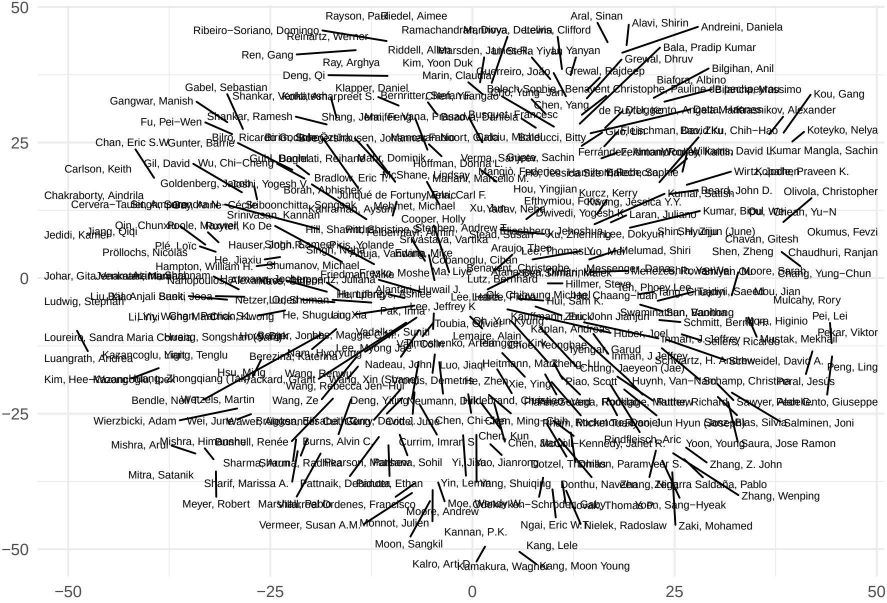


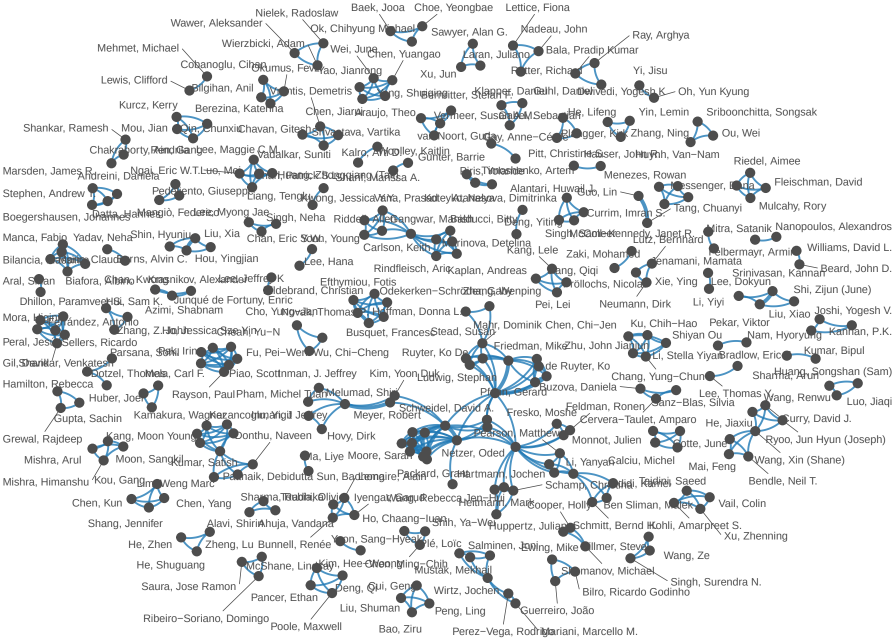
Sélection des auteurs ayant publié au moins une fois à plusieurs :
taille = nombre d'articles publiés
couleur : vert 2 et moins, coral : 3 et plus

Figure 8: network of authors

Le cas NLP et marketing

Projection Tsne des auteurs




Baek, Jooa
Choe, Yeongbae
Lettice, Fiona
Nielek, Radoslaw
Wawer, Aleksander
Ok, Chihyung Michael
Sawyer, Alan G.
Nadeau, John
Ray, Arghya
Wierzbicki, Adam
Mehmet, Michael
Wei, June
Chen, Yuangao
Laran, Juliano
Bala, Pradip Kumar
Cobanoglu, Cihan
Okumus, Fevzi
Yao, Jianrong
Xu, Jun
Ritter, Richard
Yi, Jisu
Lewis, Clifford
Bilgihan, Anil
Vrontis, Demetris
Klapper, Daniel
Dwivedi, Yogesh K
Oh, Yun Kyung
Kurcz, Kerry
Berezina, Katerina
Chen, Jiarui
Araujo, Theo
Vermeer, Susan
He, Lifeng
Yin, Lemin
Sriboonchitta, Songsak
Shankar, Ramesh
Mou, Jian
Qin, Chunxiu
Chavan, Gitesh
Srivastava, Vartika
van Noort, Guda
Gay, Anne-Céline
Zhang, Ning
Ou, Wei
Chakraborty, Arindam
Lee, Maggie C.M.
Vadalkar, Suniti
Kalro, Arpita D.
Woolley, Kaitlin
Gunter, Barrie
Pitt, Christine
Huynh, Van-Nam
Riedel, Aimee
Marsden, James R.
Andreini, Daniela
Ngai, Eric W.T.
Luo, Mei
Ilhan, Behice
Zhang, Shulin
Sharp, Marissa A.
Paris, Yolande
Timoshenko, Artem
Menezes, Rowan
Messenger, Dana
Fleischman, David
Stephen, Andrew T.
Pedeliento, Giuseppe
Liang, Tengfei
Kwong, Jessica Y.Y.
Yarlagadda, Prasad K.
Koteyko, Nelya
Alantari, Huwail J.
Guo, Lin
Tang, Chuanyi
Mulcahy, Rory
Boegershausen, Johannes
Datta, Hannes
Mangiò, Federico
Lee, Myong Jae
Singh, Neha
Riddell, Allen
Gangwar, Manish
Blasolucci, Billy
Deng, Yiting
Currim, Imran S.
Manca, Fabio
Yadav, Neha
Shin, Hyunju
Liu, Xia
Chan, Eric Y.
Son, Young
Carlson, Keith
Marinova, Detelina
Singh, Sonal
Kennedy, Janet R.
Mitra, Satanik
Nanopoulos, Alexandros
Kang, Lele
Lutz, Bernhard
Zaki, Mohamed
Felbermayr, Armin
Williams, David L.
Bilancia, Massimo
Claude
Burns, Alvin C.
Hou, Yingjian
Lee, Hana
Benamani, Mamata
Srinivasan, Kannan
Beard, John D.
Aral, Sinan
Biafora, Albino
Chan, Kwong
Krasnikov, Alexander
Lee, Jeffrey K
Rindfleisch, Aric
Kaplan, Andreas
Wang, Qiqi
Kröllochs, Nicolas
Xie, Ying
Lee, Dokyun
Shi, Zijun (June)
Efthymiou, Fotis
Hildebrand, Christian
Odekerken-Schröder, Gaby
Zhang, Wenping
Dhillon, Paramveer S.
Hsu, Sam K.
Junqué de Fortuny, Enric
Pei, Lei
Neumann, Dirk
Li, Yiyi
Azimi, Shabnam
Cho, Yunjae
Novak, Thomas P.
Hoffman, Donna L.
Liu, Xiao
Joshi, Yogesh V.
Mora, Higinio
Fernández, Antonio
Zhang, Z. John
Ho, Jessica
Sze Yin
Chean, Yu-N
Mahr, Dominik
Chen, Chi-Jen
Ku, Chih-Hao
Pekar, Viktor
Kannan, P.K.
Peral, Jesús
Sellers, Ricardo
Busquet, Francesc
Stead, Susan
Shiyan Ou
Nam, Hyoryung
Parsana, Sohil
Paki, Irina
Fu, Pei-Wen
Wu, Chi-Cheng
Ruyter, Ko De
Friedman, Mike
Zhu, John Jianjun
Li, Stella Yiyan
Bradlow, Eric T.
Kumar, Bipul
Gil, David
Shankar, Venkatesh
Dotzel, Thomas
Mela, Carl F.
Piao, Scott
Inman, J. Jeffrey
Kim, Yoon Duk
Ludwig, Stephan
de Ruyter, Ko
Chang, Yung-Chun
Sharma, Arun
Huang, Songshan (Sam)
Hamilton, Rebecca
Rayson, Paul
Pham, Michel Tuan
Melumad, Shiri
Pfann, Gerard
Buzova, Daniela
Lee, Thomas Y.
Wang, Renwu
Luo, Jiaqi
Huber, Joel
Schweidel, David A.
Feldman, Ronen
Sanz-Blas, Silvia
Gupta, Sachin
Kamakura, Wagner
Kazancoglu, Ipek
Inman, J. Jeffrey
Meyer, Robert
Fresko, Moshe
Cervera-Taulet, Amparo
He, Jiaxiu
Curry, David J.
Grewal, Rajdeep
Kang, Moon Young
Donthu, Naveen
Hovy, Dirk
Pearson, Matthew
Monnot, Julien
Cotte, June
Ryoo, Jun Hyun (Joseph)
Mishra, Arul
Moon, Sangkil
Kumar, Satish
Ma, Liye
Moore, Sarah
Netzer, Oded
Li, Yanyan
Calciu, Michel
Mai, Feng
Wang, Xin (Shane)
Mishra, Himanshu
Kou, Gang
Lim, Weng Marc
Pattnaik, Debidutta
Sun, Baohong
Lemaire, Alain
Packard, Grant
Hartmann, Jochen
Schamp, Christina
Teidini, Kamel
Taidini, Saeed
Bendle, Neil T.
Chen, Kun
Chen, Yang
Sharma, Rabindra
Toubia, Olivier
Iyengar, Garud
Wang, Rebecca Jen-Hui
Heitmann, Mark
Cooper, Holly
Ben Sliman, Malek
Vail, Colin
Shang, Jennifer
Alavi, Shirin
Ahuja, Vandana
Ho, Chaang-Iuan
Shih, Ya-Wei
Huppertz, Julian
Schmitt, Bernd H.
Kohli, Amarpreet S.
Xu, Zhenning
He, Zhen
Zheng, Lu
Bunnell, Renée
Yoon, Sang-Hyeak
Doré, Loïc
Salminen, Joni
Ewing, Mike
Hillmer, Steven
Wang, Ze
He, Shuguang
McShane, Lindsay
Kim, Hee-Woong
Chen, Ming-Chih
Mustak, Mekhail
Shumanov, Michael
Singh, Surendra N.
Saura, Jose Ramon
Deng, Qi
Cui, Geng
Wirtz, Jochen
Bilro, Ricardo Godinho
Pancer, Ethan
Liu, Shuman
Peng, Ling
Guerreiro, João
Ribeiro-Soriano, Domingo
Poole, Maxwell
Bao, Ziru
Perez-Vega, Rodrigo
Mariani, Marcello M.

L’exemple d’un réseau de citation

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)
- Tranformers is all you need, now.

The LDA model

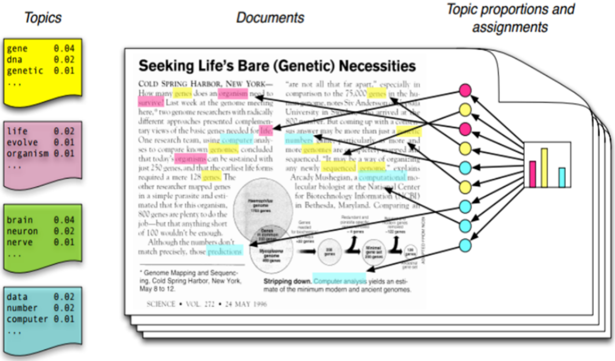
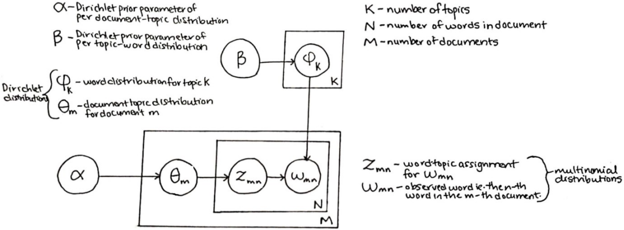


Figure 9: The LDA concept


α - Dirichlet prior parameter of per document-topic distribution
β - Dirichlet prior parameter of per topic-word distribution
Dirrchlet distribution
φ_k - word distribution for topic k
θ_m - document topic distribution for document m
K - number of topics
N - number of words in document
M - number of documents
β
φ_k
K
α
θ_m
z_{mn}
w_{mn}
N
M
z_{mn} - word topic assignment for w_{mn}
w_{mn} - observed word ie. the n-th word in the m-th document.
multinomial distributions

Now a large family with Structural Topic Models or Seed LDA models.

An application of a STM model

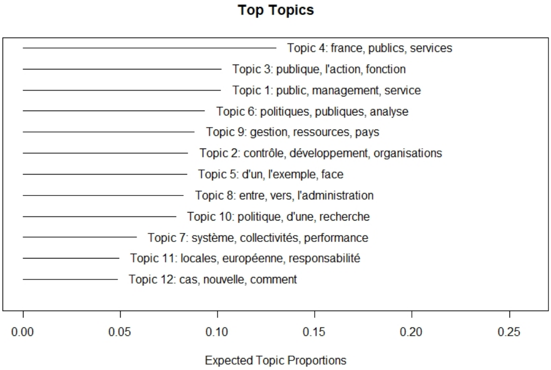


Figure 10: A STM topic model : keywords and proportion of the content /n (each document has a p probability to belong to the topic k)

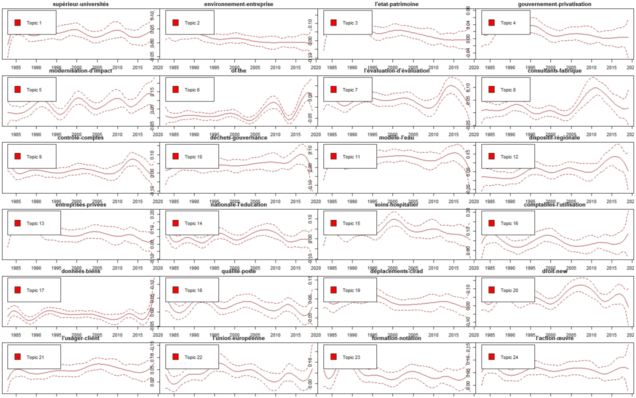


Figure 11: A STM topic model : time prevalence for each topic identified

An embeddings approach

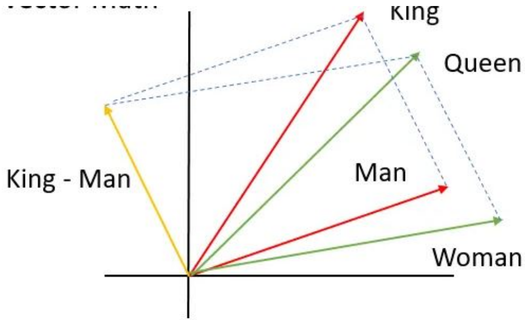


Figure 12: Embeddings intuition

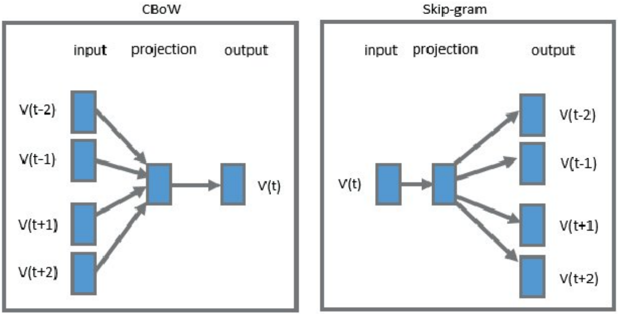


Figure 13: The word2vec model

Annotation stage

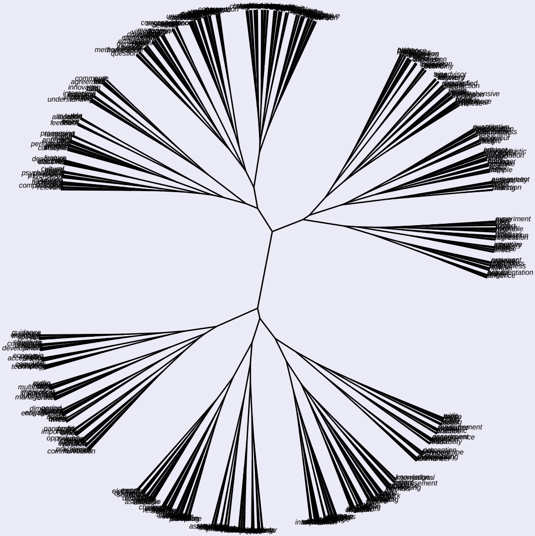
```
2023-02-16 08:33:40 Annotating text fragment 1/110
2023-02-16 08:33:46 Annotating text fragment 21/110
2023-02-16 08:33:54 Annotating text fragment 41/110
2023-02-16 08:34:02 Annotating text fragment 61/110
2023-02-16 08:34:09 Annotating text fragment 81/110
2023-02-16 08:34:18 Annotating text fragment 101/110
Time difference of 42.73675 secs
```

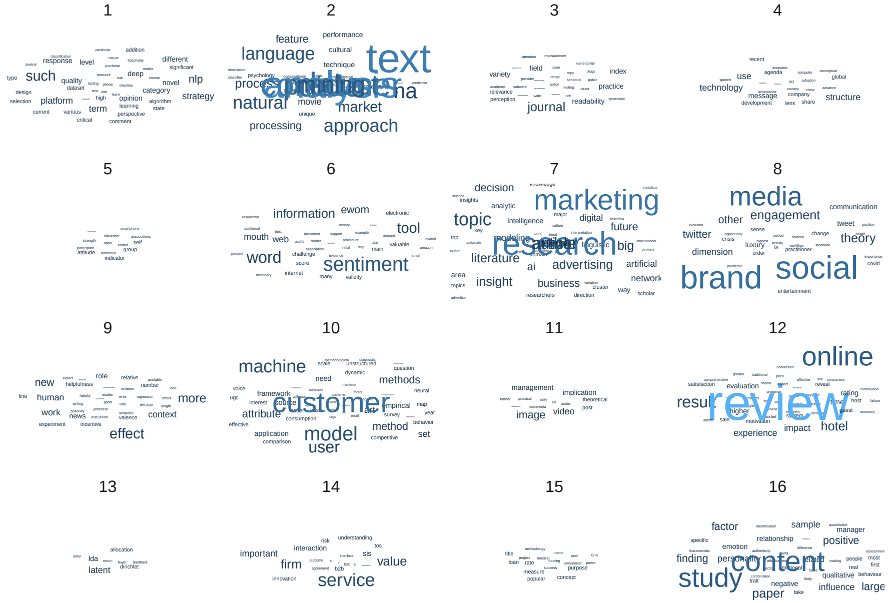
vectorisation Stage

Un test

```
$lda
   term1         term2 similarity rank
1    lda     dirichlet  0.8190907    1
2    lda    allocation  0.8050508    2
3    lda        latent  0.6876944    3
4    lda       mindset  0.6287284    4
5    lda psychological  0.6183707    5
6    lda        course  0.6016998    6
7    lda         award  0.5977740    7
8    lda   description  0.5879406    8
9    lda     awareness  0.5843993    9
10   lda       chinese  0.5599142   10
11   lda         model  0.5524023   11
12   lda         topic  0.5437085   12
13   lda       outcome  0.5366894   13
14   lda         other  0.5325939   14
15   lda   methodology  0.5320520   15
16   lda       balance  0.5313188   16
17   lda    industries  0.5309694   17
18   lda      relevant  0.5299399   18
19   lda          loan  0.5275514   19
20   lda        seller  0.5235679   20
```

Word clustering Stage

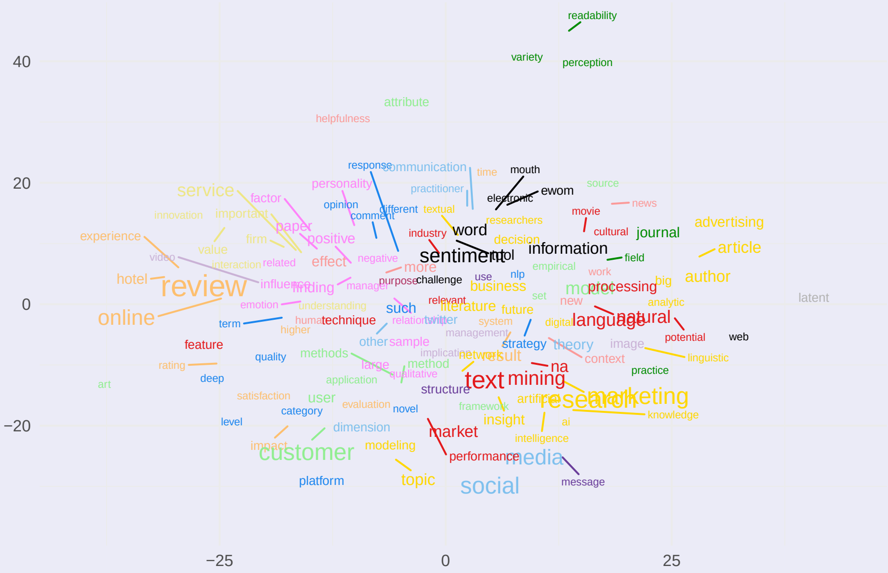



1
response
level
different
significant
addition
particular
classification
several
nature
hospitality
purchase
mobile
type
such
quality
dataset
resource
cue
deep
course
novel
nlp
design
selection
platform
testing
phone
intention
category
test
aim
high
learn
opinion
algorithm
strategy
learning
state
current
various
term
perspective
critical
comment
2
feature
performance
language
cultural
technique
text
description
classifier
psychology
computational
predictive
process
natural
movie
market
unique
processing
approach
3
attention
measurement
vulnerability
field
issue
variety
copy
blogs
index
provider
range
sensorial
public
policy
feeling
direct
practice
academic
relevance
software
wide
rich
perception
journal
readability
systematic
4
recent
economic
agenda
computer
conceptual
use
global
speech
act
adoption
technology
country
cross
acceptance
message
company
advance
development
lens
share
structure
5
smartphone
influencer
associations
strength
open
ended
self
participant
attitude
reference
group
indicator
6
researcher
information
ewom
electronic
money
tool
additional
best
document
support
example
amount
mouth
web
useful
reader
procedure
star
overall
initial
step
main
valuable
amazon
present
association
challenge
evidence
small
word
score
sentiment
dictionary
internet
many
validity
7
decision
knowledge
statistical
science
insights
analytic
marketing
topic
intelligence
major
digital
overview
key
culture
future
print
trend
interpretation
top
automate
modeling
research
linguistic
big
international
extant
literature
domain
journals
ai
advertising
artificial
area
network
insight
business
variation
cluster
topics
way
researchers
direction
scholar
advertise
8
media
communication
evolution
other
engagement
tweet
position
sense
health
twitter
opportunity
crisis
period
balance
change
theory
luxury
highest
activity
workflow
facebook
dimension
order
tv
practitioner
brand
social
importance
covid
pandemic
entertainment
9
new
expert
role
relative
available
helpfulness
reviewer
number
idea
line
human
helpful
retailer
write
regression
affect
more
good
writing
vote
diffusion
length
work
positivity
processe
news
discussion
sentence
valence
context
experiment
incentive
effect
10
machine
methodological
scale
diagnostic
unstructured
question
need
dynamic
methods
voice
marketer
neural
common
framework
focus
ugc
patterns
collector
interest
source
customer
empirical
map
attribute
art
survey
year
consumption
tags
retail
behavior
effective
method
application
model
competitive
set
comparison
user
11
management
implication
further
practical
selfy
ud
audio
theoretical
multimedia
post
image
video
12
online
conversion
greater
traditional
price
comprehensive
affective
site
consumers
satisfaction
evaluation
theme
aspect
reveal
presence
contribution
result
review
rating
host
failure
higher
guest
economy
airbnb
sale
motivation
recovery
system
experience
impact
hotel
13
allocation
seller
lda
return
buyer
feedback
dirichlet
latent
14
risk
understanding
tos
important
interaction
sis
interface
firm
outcome
si
value
b2c
agreement
b2b
innovation
service
15
methodology
metric
title
peer
form
project
mindset
loan
rate
funding
awareness
power
success
purpose
measure
popular
concept
16
factor
identification
sample
quantitative
manager
relationship
specific
positive
emotion
authenticity
difference
characteristic
assessment
finding
personality
content
related
making
people
most
real
first
study
combination
qualitative
behaviour
trait
likely
negative
influence
large
paper
fake

projections Tsne

Projection 2D du vocabulaire vectorisé des 110 articles 'NLP & Marketing'

Pipe : annot. syntax −> word2vec: 200 vecteurs−> hclus −>rtsne

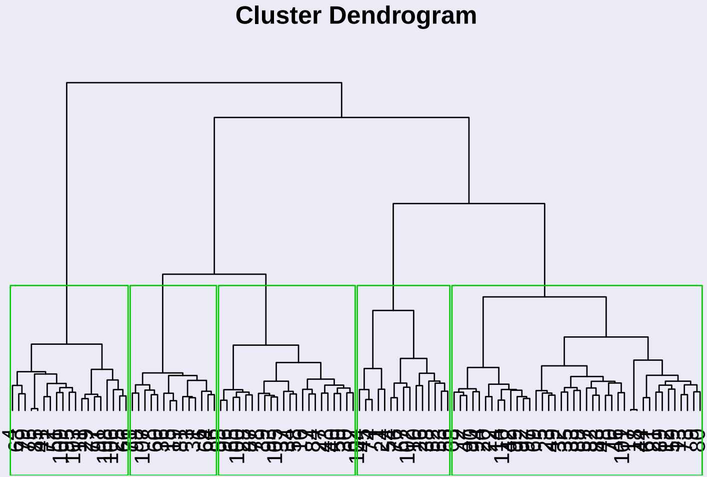


document embeddings

Text vectors are created simply by calculating the resulting vector of words in the text :
text = *le NLP quantifie les mots*
$V_text =V_nlp+ V_quantifier+ V_mots$
We could also compute abstract concepts. for exemple, finding mention of advanced NLP methods : - concept="LDA Word2vec Bert ML" - compute the similarity between each text and the concept

document clustering

grouping documents

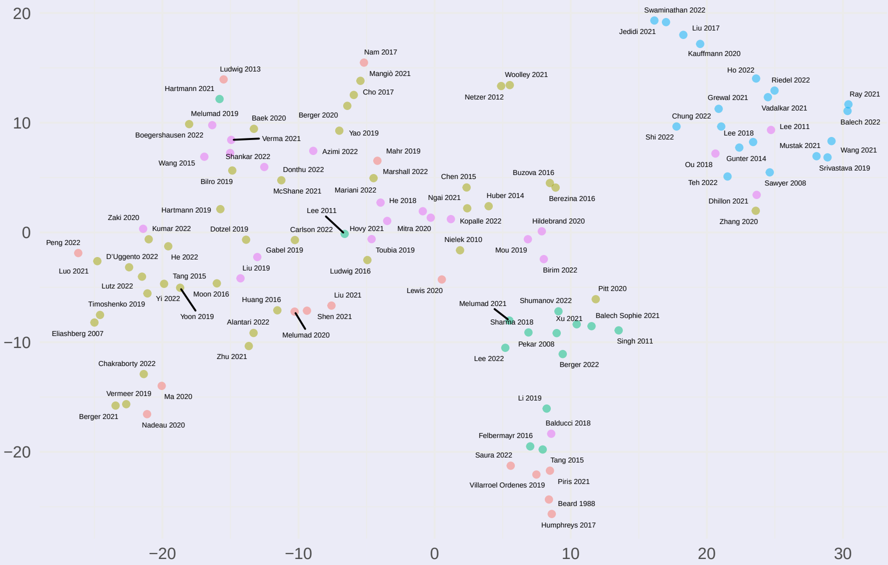


document Tsne

document projection space

Projection 2D des 106 articles 'NLP & Marketing' vectorisés

Pipe : annot. syntax −> word2vec: 200 vecteurs−> doc2vec−> hclus −>rtsne

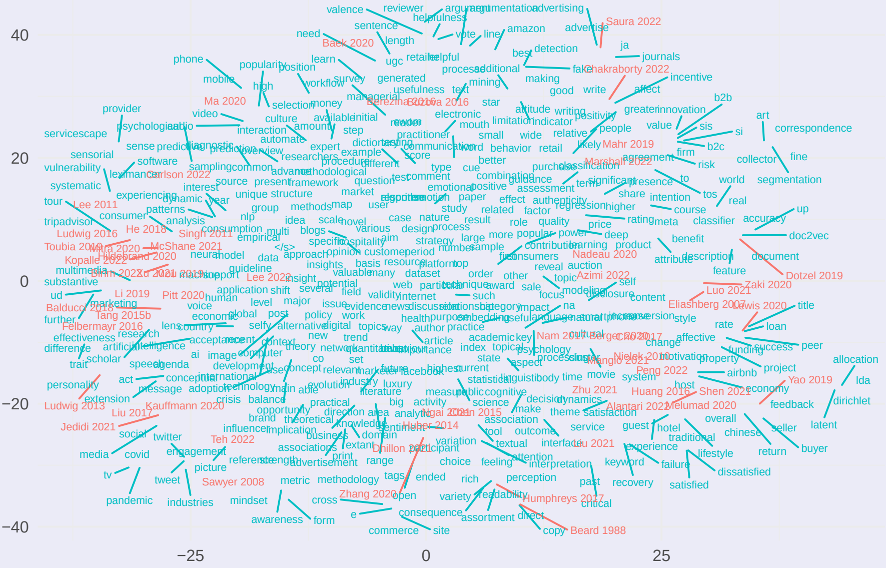


We might as well put everything in one space

Because embedding space is common.

Projection 2D du vocabulaire vectorisé des 110 articles 'NLP & Marketing'

Pipe : annot. syntax −> word2vec: 200 vecteurs−> hclus −>rtsne



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

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.