

Topic Modeling Classification of Publications on CSR and Technology Transfer

Moses A. Boudourides

Professor of Practice
School of Public Affairs
Arizona State University

Moses.Boudourides@asu.edu

Draft Report
July 2022

The dataset

- ▶ On June 6, 2022, the **Web of Science (WoS)** database was searched by selecting the following keywords and Boolean commands:
 - ▶ *Title*: (“technology transfer” OR “corporate social responsibility”) AND (“management” OR “public policy”) OR
 - ▶ *Abstract*: (“entrepreneur” OR “firm” OR “responsible innovation”)
- ▶ Thus, 529 records (**publications** or **articles**) were retrieved, attributed in 70 fields (columns), among which (11 are shown in a sample of the dataset in next slide) were retained for the present study.

A sample of the dataset

	Type	Title	Source	Authors	Editors	Year	Author Keywords	Keywords Plus	Abstract	Times Cited, WoS Core	Times Cited, All Databases
481	J	Measuring systematic-risk using implicit beta	Manage. Sci.	SIEGEL, AF	NaN	1995	EXCHANGE OPTION; VOLATILITY AND RETURN IN FINA...	NaN	A new technology is proposed for estimating th...	18	18
487	J	Corporate social responsibility and financial ...	Strateg. Manage. J.	McWilliams, A; Siegel, D	NaN	2000	corporate social responsibility; firm performa...	R-AND-D; INDUSTRY MATTER; INVESTMENT; EVENT; P...	Researchers have reported a positive, negative...	1739	1775
8	J	Corporate social responsibility and management...	J. Bus. Ethics	Lee, D	NaN	2017	Corporate social responsibility; Disclosure re...	SARBANES-OXLEY ACT; STAKEHOLDER MANAGEMENT; EA...	This study examines the association between co...	32	32
355	J	Management of technology-transfer - can it be ...	NaN	DEBETTIGNIES, HC	NaN	1978	NaN	NaN	NaN	0	0
468	J	Lessons learned for a more efficient knowledge...	Waste Manage. Res.	Bezama, A; Szarka, N; Navia, R; Konrad, O; Lor...	NaN	2007	municipal solid waste management; industrial w...	TANNERY	The present paper describes the development, p...	5	5
510	J	Issues in the use of the event study methodolo...	Organ. Res. Methods	McWilliams, A; Siegel, D; Teoh, SH	NaN	1999	NaN	NaN	Organizational researchers are increasingly us...	67	68

Motivation

- ▶ In *bibliographic social network analysis*, besides the common cases of citation (co-citation and bibliographic coupling) and co-authorship networks, some authors have also studied **keyword co-occurrence networks** (Maltseva & Batagelj [2018, 2019], Leydesdorff et al. [2008], Groenewegen et al. [2015]).
- ▶ Typically, WoS catalogues two types of keywords for each archived article (called *publication* from now on):
 - ▶ **Author Keywords**: chosen by the author to best reflect the content of the document, and
 - ▶ **Keywords Plus**: index terms automatically generated from the titles of cited articles.
- ▶ Instead of these, here, we are using a *Machine Learning* and *Natural Language Processing* approach in order to assign the hidden semantic structures of a text body, discovered by a **Topic Model**, as another type of keywords, which are endogenously extracted from the text (or abstract) of publications (independently of exogenous allocations made by authors or archivers).

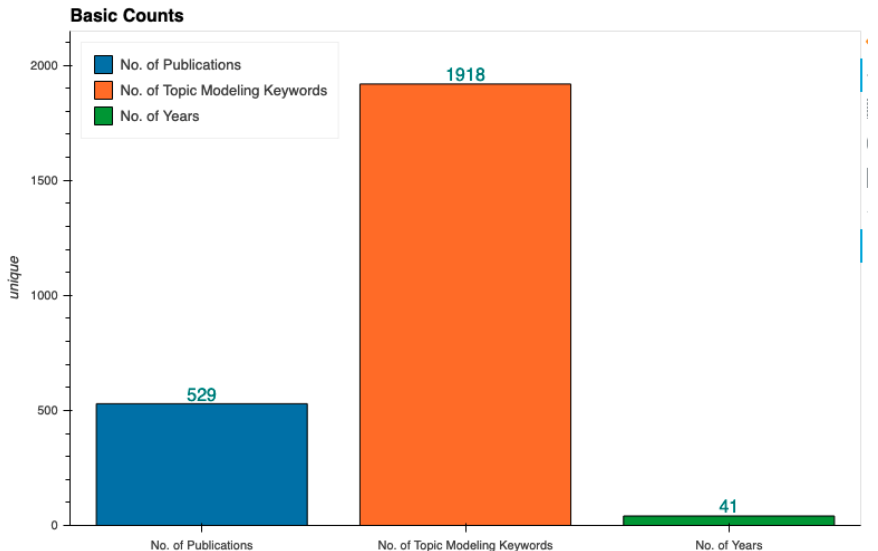
Topic Modeling Classification, I

- ▶ Suppose that we have a collection of documents, each one being composed of nontrivial words or phrases.
- ▶ Here, each document is a publication and we are only considering words in the abstract. *However, in a future furthering of our study, we intend to consider the whole text of each publication.*
- ▶ **Topic Modeling** is an unsupervised machine learning technique that proceeds in two stages.
- ▶ In the first stage, after scanning the set of documents, Topic Modeling produces a *vocabulary of pre-processed* (“cleaned,” uncapitalized, lemmatized etc.) words (or phrases).
- ▶ Here, these words will be called **Topic Modeling keywords**.

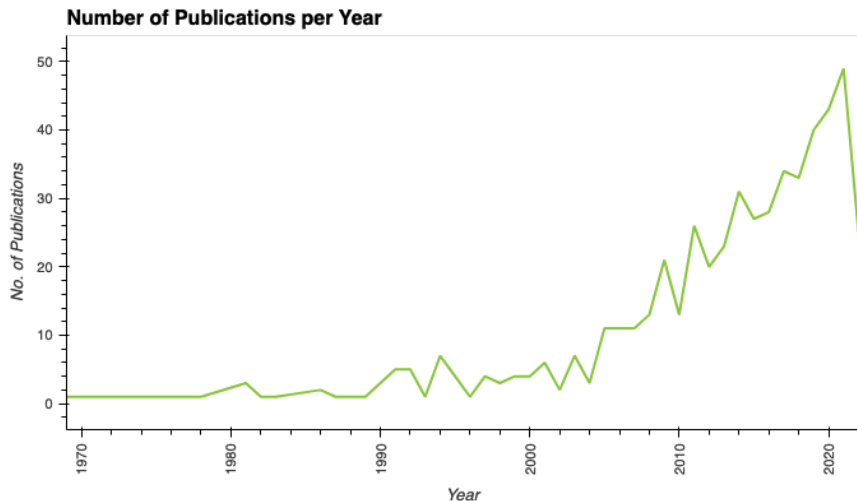
Topic Modeling Classification, II

- ▶ In the second stage, using **Latent Dirichlet Allocation (LDA)**, which is a generative statistical model, what Topic Modeling does is the following:
- ▶ Instead of representing a document in its feature space (by considering frequencies of words in each document), it clusters (classifies) all the words of the vocabulary in a *topic space*, consisting of a given number of *Topics*, i.e., groups of words that are associated under a single interpretable theme, and it assess two types of probabilities:
 - ▶ the weights of words in each document to be assigned to each topic, and
 - ▶ the strength with which each document exhibits each one of the topics.
- ▶ Thus, according to the weights that words in a document possess in order to be assigned to each topic, for each document, there exists a **dominant topic** and a corresponding **probability contribution** for the document to be assigned to a dominant topic.

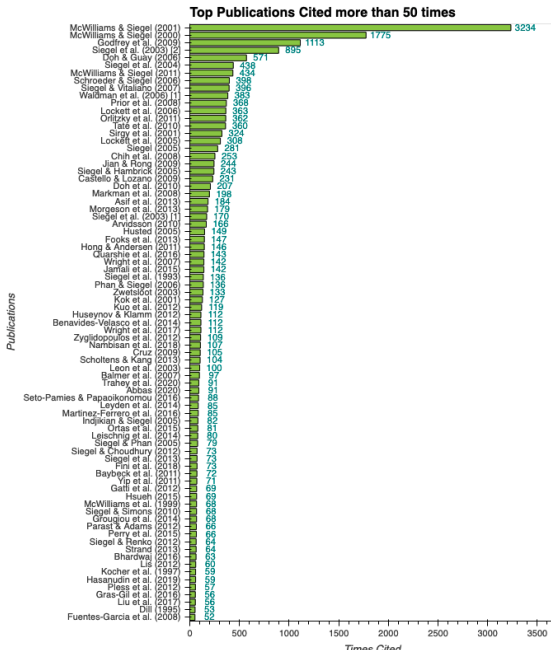
The dataset of publications from 1969 to 2022



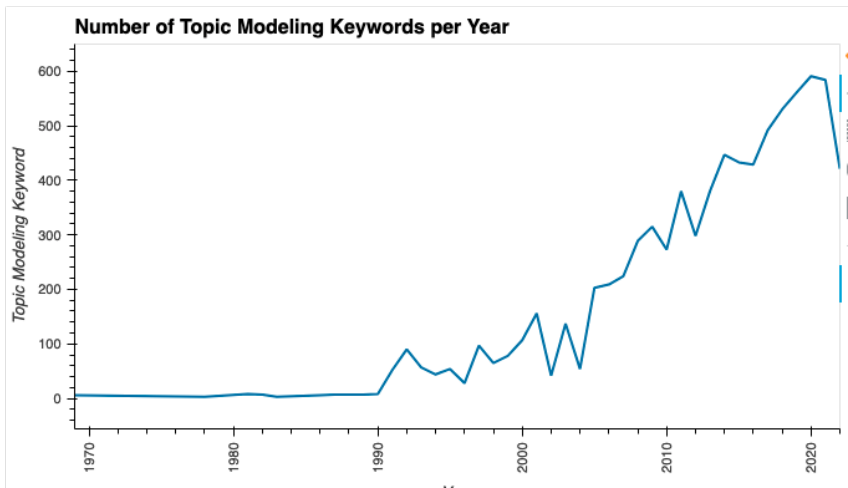
Publications per Year



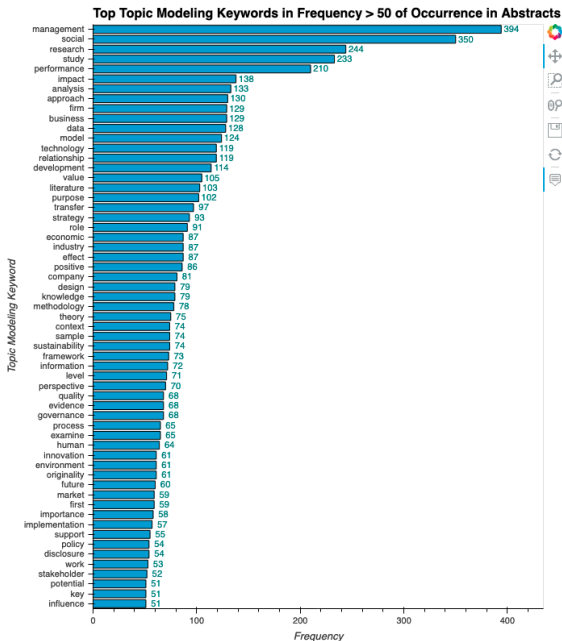
Top Publications in Times Cited



Topic Modeling Keywords per Year

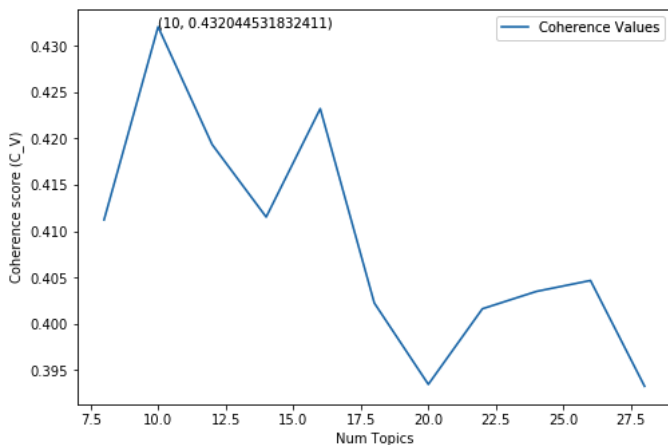


Top Topic Modeling Keywords



Why 10 Topics?

Maximization of the CV coherence score



Wordclouds of Topics and Topic Modeling Keywords

Topic Modeling of the Web of Science bibliographic dataset on CSR and Technology Transfer

Topic 1 (Governance)
bonding methodology law
governance
criterion fee awareness
environment tax battery
engagement forecast
accounting listing
evidence security
sophistication action
firm energy storage

Topic 4 (Management Models)
development case
analysis
system project
program transfer waste
management
country
technology research paper
knowledge process government study
model innovation
product

Topic 7 (Work Organization)
policy organization
human_resource attitude
behavior performance
employee work
identification
study commitment
support practice
firm worker
change management job
engagement workforce

Topic 10 (Firm Performance)
responsibility
capital consumer
family market
management value
relationship
performance theory
risk policy engagement
impression ownership image
stakeholder
firmEffect brand

Topic 2 (Management Studies)
effect practice research
management impact
firm stakeholder risk
responsibility approach
relationship finding
disclosure activity
performance
company study
earning result quality

Topic 5 (Managerial Attributes)
team market
self management
delegation
buyer health idea
education strategy
shell age
law institution
medium labor level
tiertechnology study

Topic 8 (Knowledge Productivity)
retailer
portfolio student
knowledge
result project
centric
management recognition
prostate marketing market
relationship firm opportunity
study technology transfer
capability technology
finding

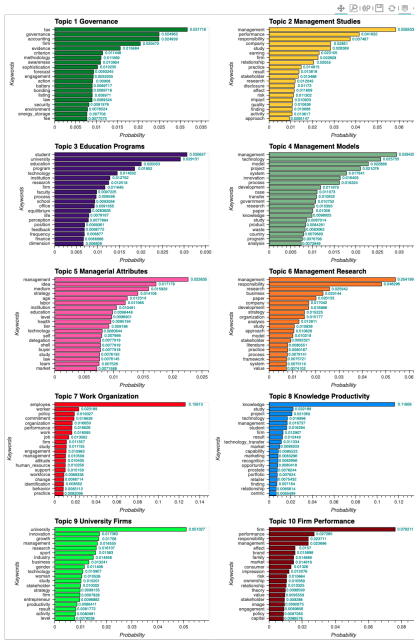
Topic 3 (Education Programs)
research feedback position
faculty student
life
firm technology frequency
equilibrium perception
education school
program office
process institution
finance
university
dimension

Topic 6 (Management Research)
management
organization value
research approach
framework stakeholder
development literature
process company
paper strategy study
system analysis
responsibility
practice business model

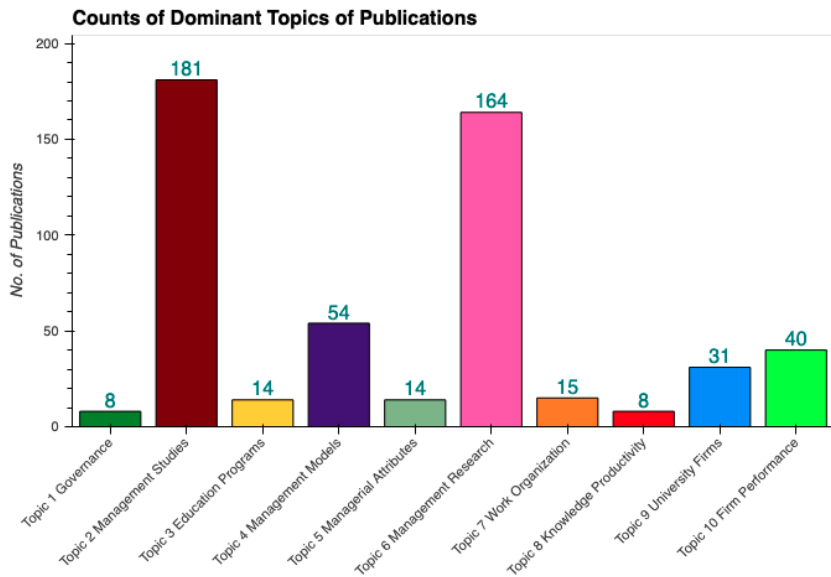
Topic 9 (University Firms)
business gender
strategy
innovation
industry stakeholder
technology growth study
woman result productivity
entrepreneur
university
management level
sport firm research
activity

Weights of Topic Modeling Keywords

Topic Modeling Keywords in Topics

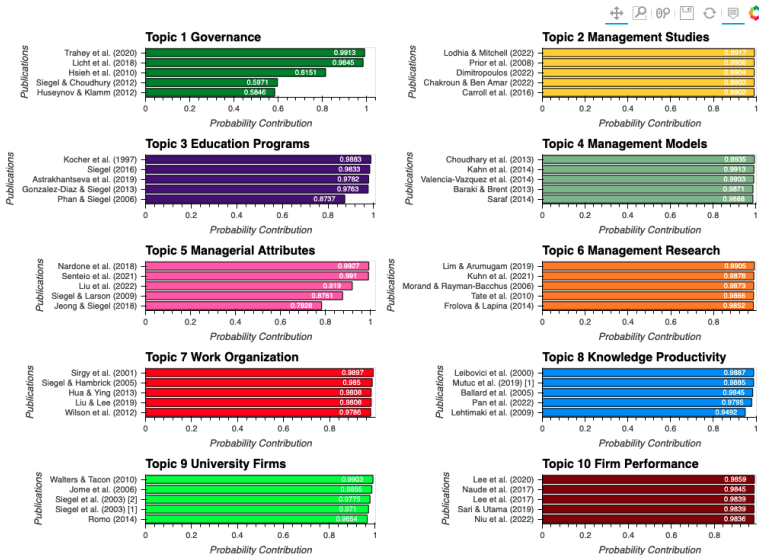


Dominant Topics of Publications



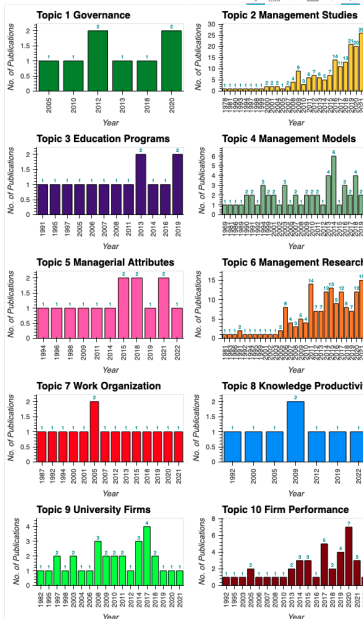
Top Dominant Topic Publications in Probability Contribution

Top Dominant Topic Publications in Probability Contribution



Number of Publications within Dominant Topics per Year

Number of Publications within Dominant Topics per Year



Top Dominant Topic Publications in Times Cited

Top Dominant Topic Publications in Times Cited

