Topic Modeling Classification of Publications on Anticipatory Governance

Moses A. Boudourides

Professor of Practice
School of Public Affairs
Arizona State University
Moses.Boudourides@asu.edu

Draft Report

August 2022

The dataset

- On August 17, 2022, the **Web of Science** (**WoS**) database was searched by selecting the following keyword:
  - *All Fields*: ("`Anticipatory Governance`")
- Thus, 168 records (**publications** or **articles**) were retrieved, attributed in 70 fields (columns), among which 86 records were including the Abstract field and were retained for the present study (the next slide displays a sample of the dataset including 12 columns).

A sample of the dataset

	Type	Title	Source	Authors	Editors	Year	Author Keywords	Keywords Plus	Abstract	Times Cited, WoS Core	Times Cited, All Databases	WoS Categories
55	J	Nanotechnology for sustainability: what does n...	J. Nanopart. Res.	Wiek, A; Foley, RW; Guston, DH	NaN	2012	Nanotechnology; Sustainability; Complex proble...	UNITED-STATES; SCIENCE; POLICY; GOVERNANCE; HE...	Nanotechnology is widely associated with the p...	24	24	Chemistry, Multidisciplinary; Nanoscience & Na...
10	J	Nanotechnology in the city: sustainability cha...	J. Urban Technol.	Wiek, A; Guston, D; van der Leeuw, S; Selin, C...	NaN	2013	Nanotechnology; City; Urban Visions; Sustainab...	EMERGING TECHNOLOGIES; RISK PERCEPTIONS; SCIEN...	Visions about the use of nanotechnologies in t...	22	22	Urban Studies
65	J	Broken promises and breaking ground for respon...	Technol. Anal. Strateg. Manage.	Wiek, A; Foley, RW; Guston, DH; Bernstein, MJ	NaN	2016	Responsible innovation; sustainability; interv...	ANTICIPATORY GOVERNANCE; SCIENCE; NANOMATERIAL...	Despite repeated calls for novel forms of inno...	13	13	Management; Multidisciplinary Sciences
48	J	'you caught me off guard': probing the futures...	J. Nanopart. Res.	Sadowski, J; Guston, DH	NaN	2016	Complex engineered nanomaterials; Technology a...	NANOTECHNOLOGY; SOCIETY; SCIENCE	This paper applies principles and methods from...	5	5	Chemistry, Multidisciplinary; Nanoscience & Na...
110	J	A roadmap for gene drives: using institutional...	J. Responsible Innov.	Kuzma, J; Gould, F; Brown, Z; Collins, J; Delb...	NaN	2018	Gene drive; governance; risk; systems; IAD; ge...	PRECAUTION; THINKING	The deployment of gene drives is emerging as a...	25	25	Ethics; History & Philosophy Of Science; Manag...

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 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 hiden 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).

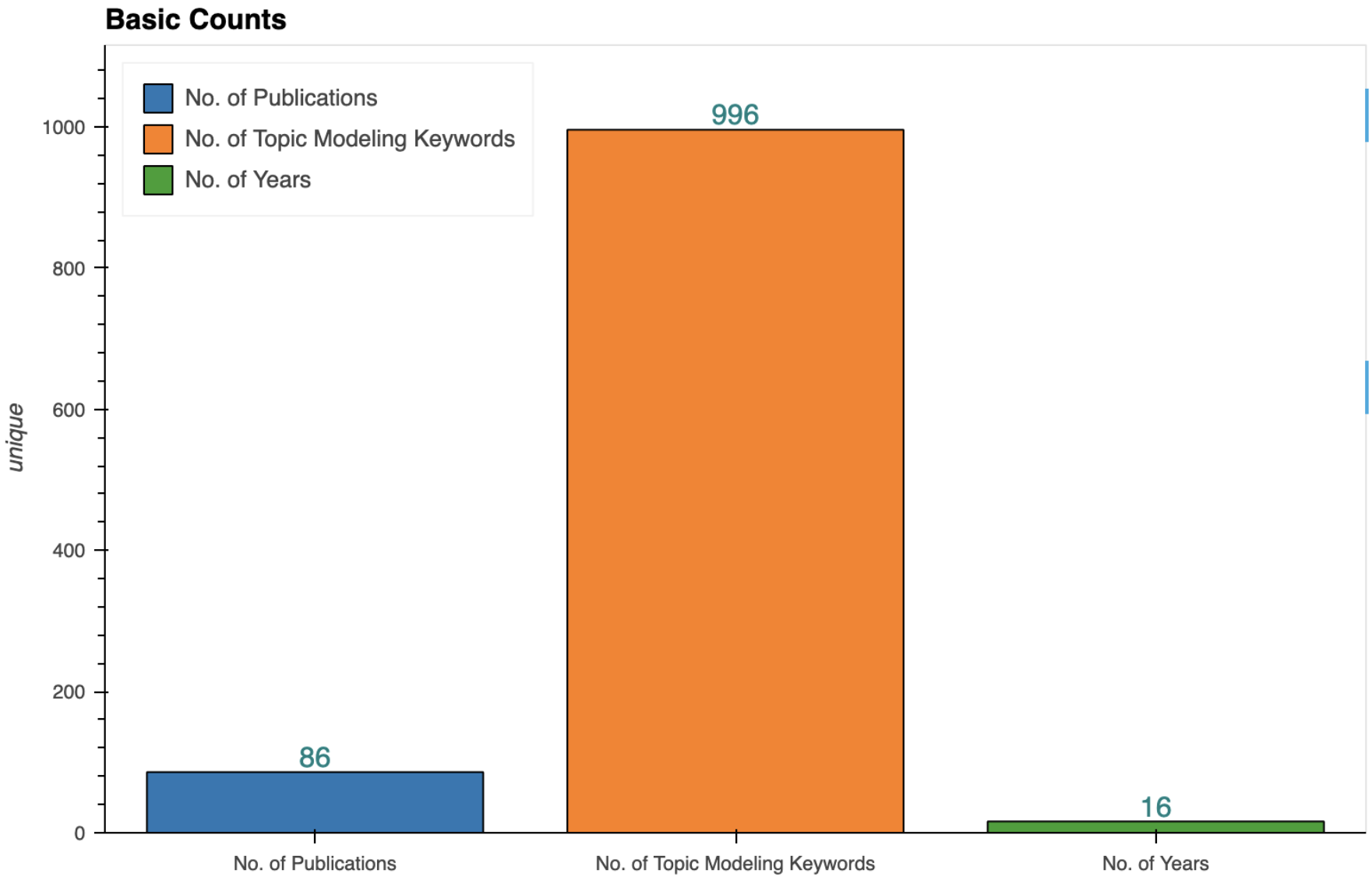
Topic Modeling Classification, I

- 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**.

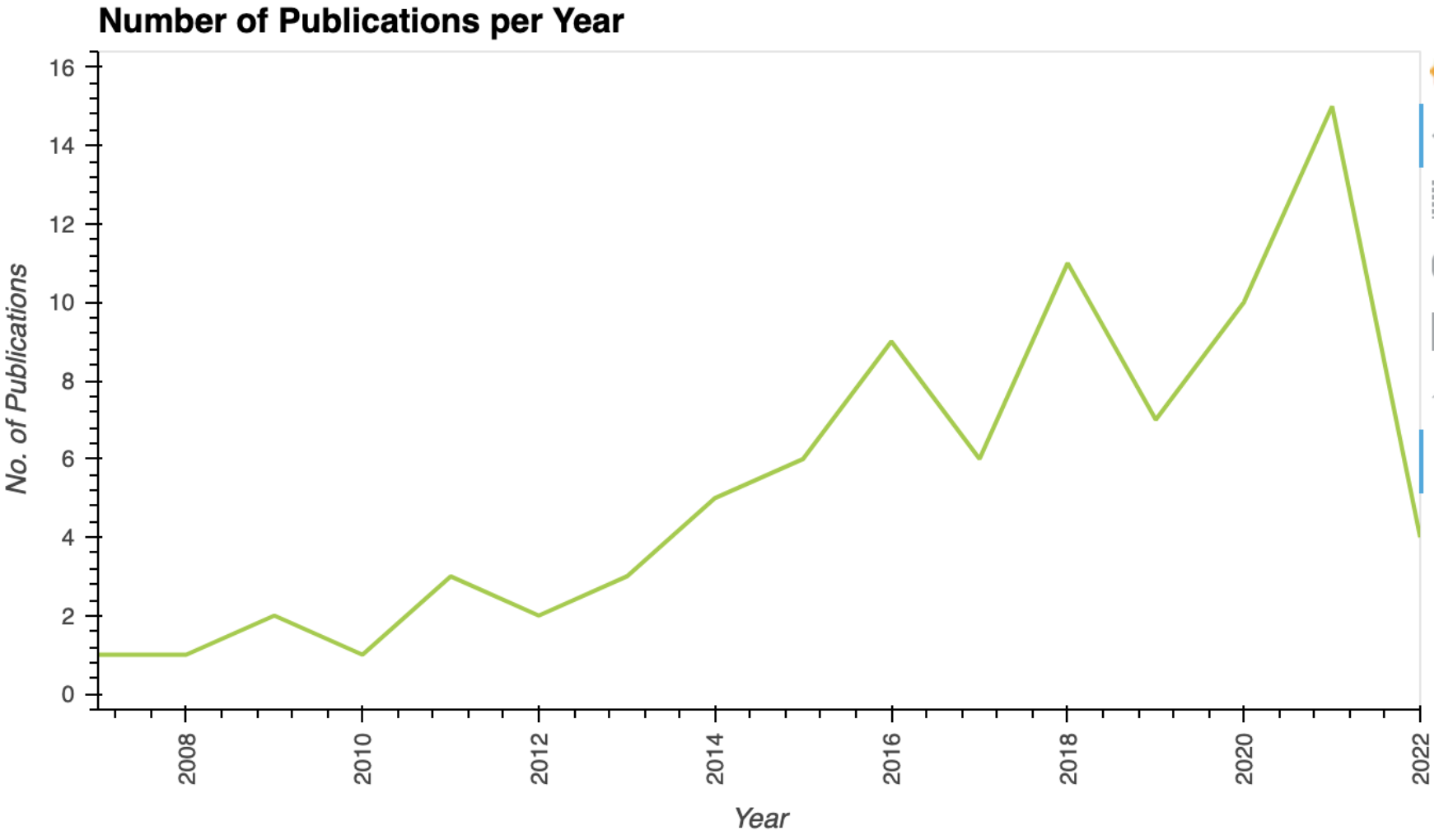
Topic Modeling Classification, II

- 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.

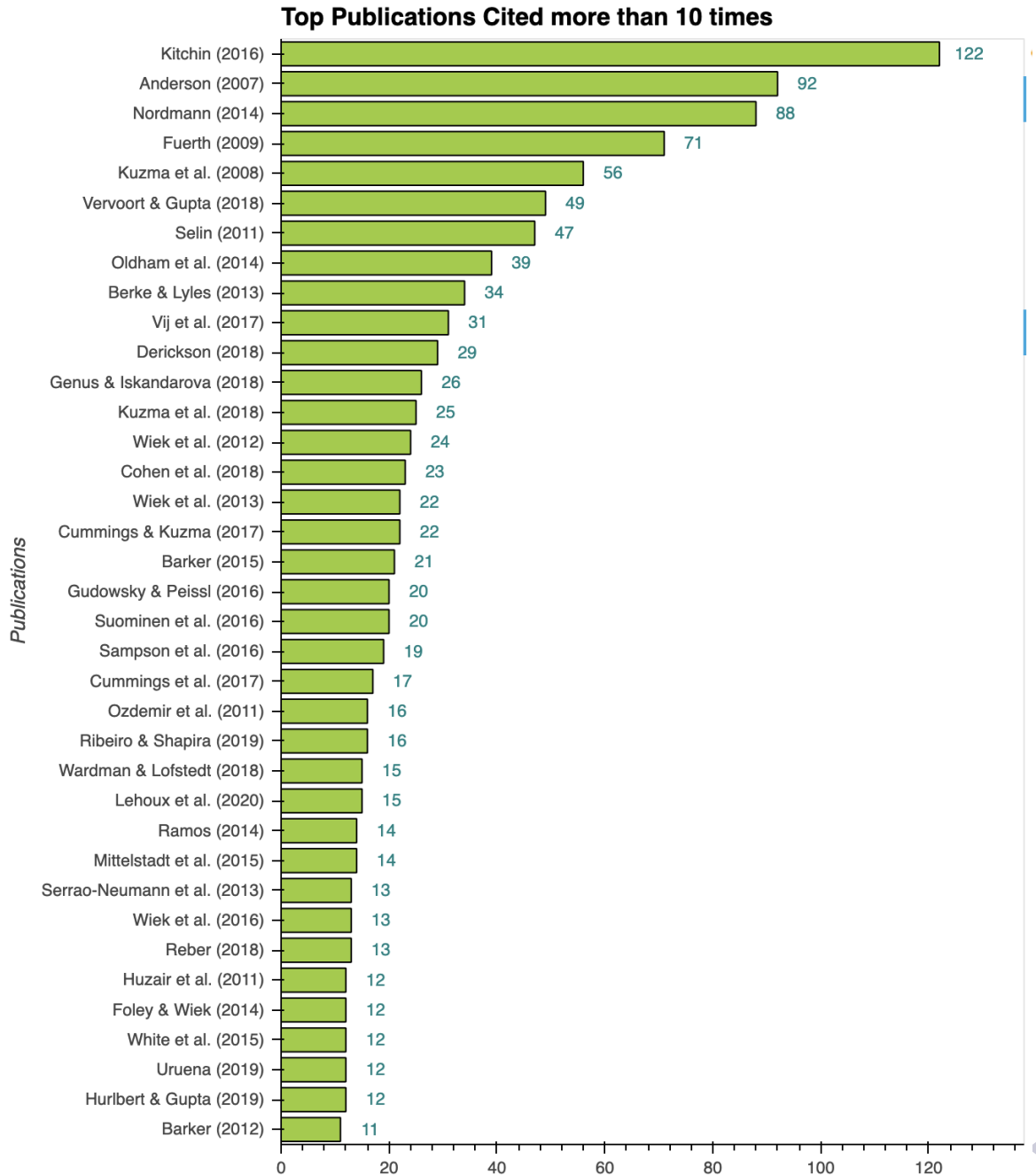
The dataset of publications from 1965 t0 2022



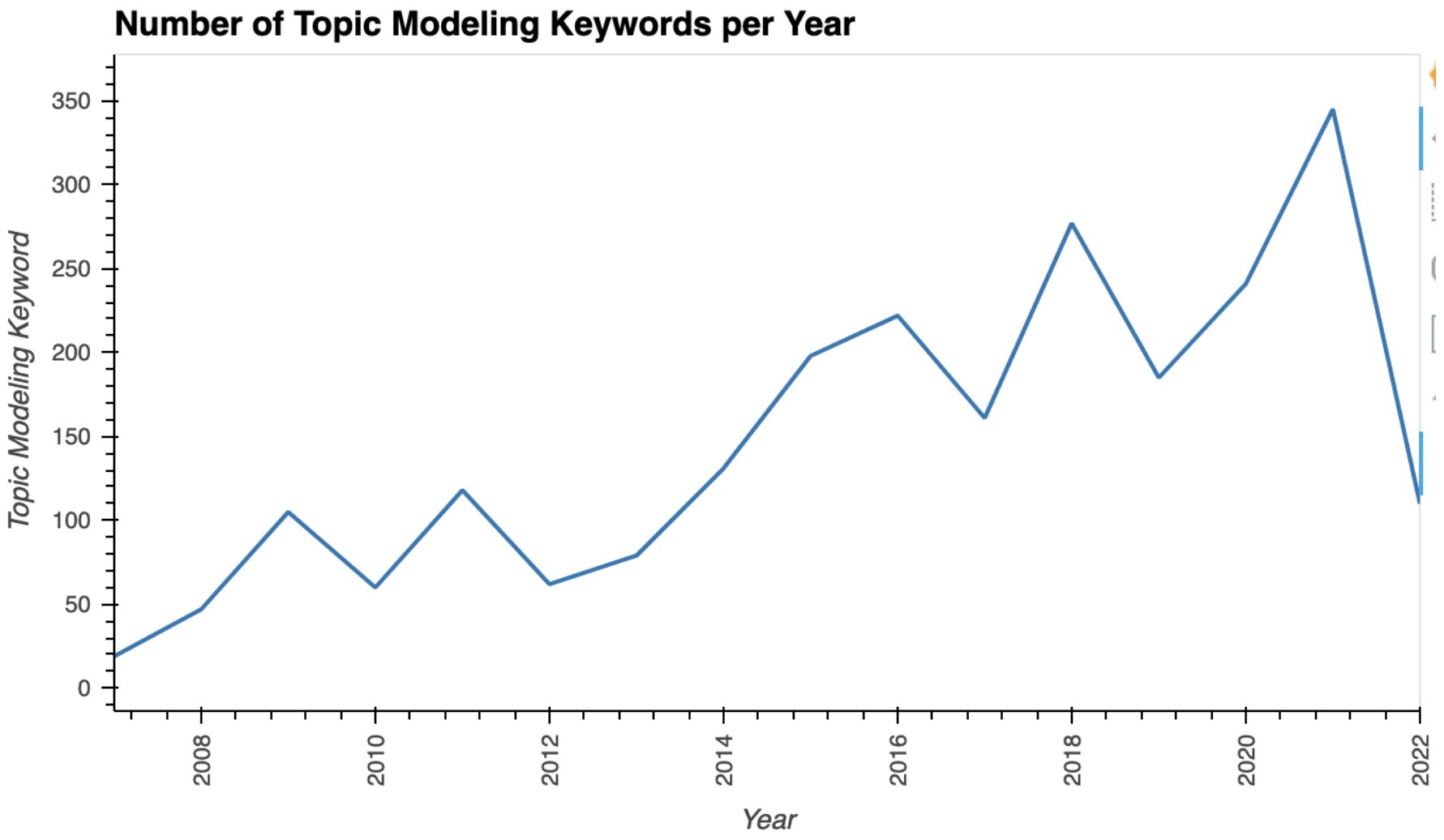
Publications per Year



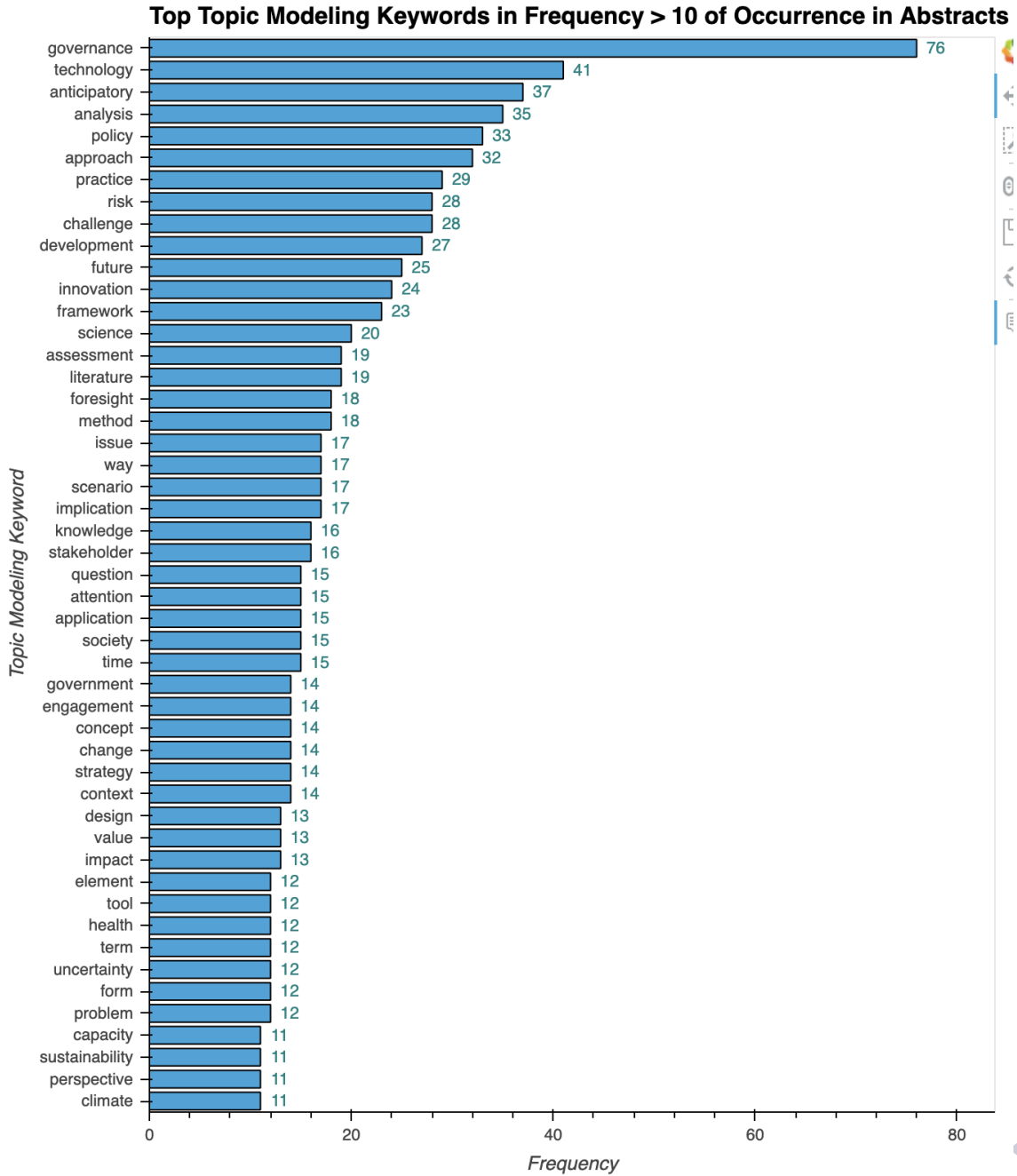
Top Publications in Times Cited



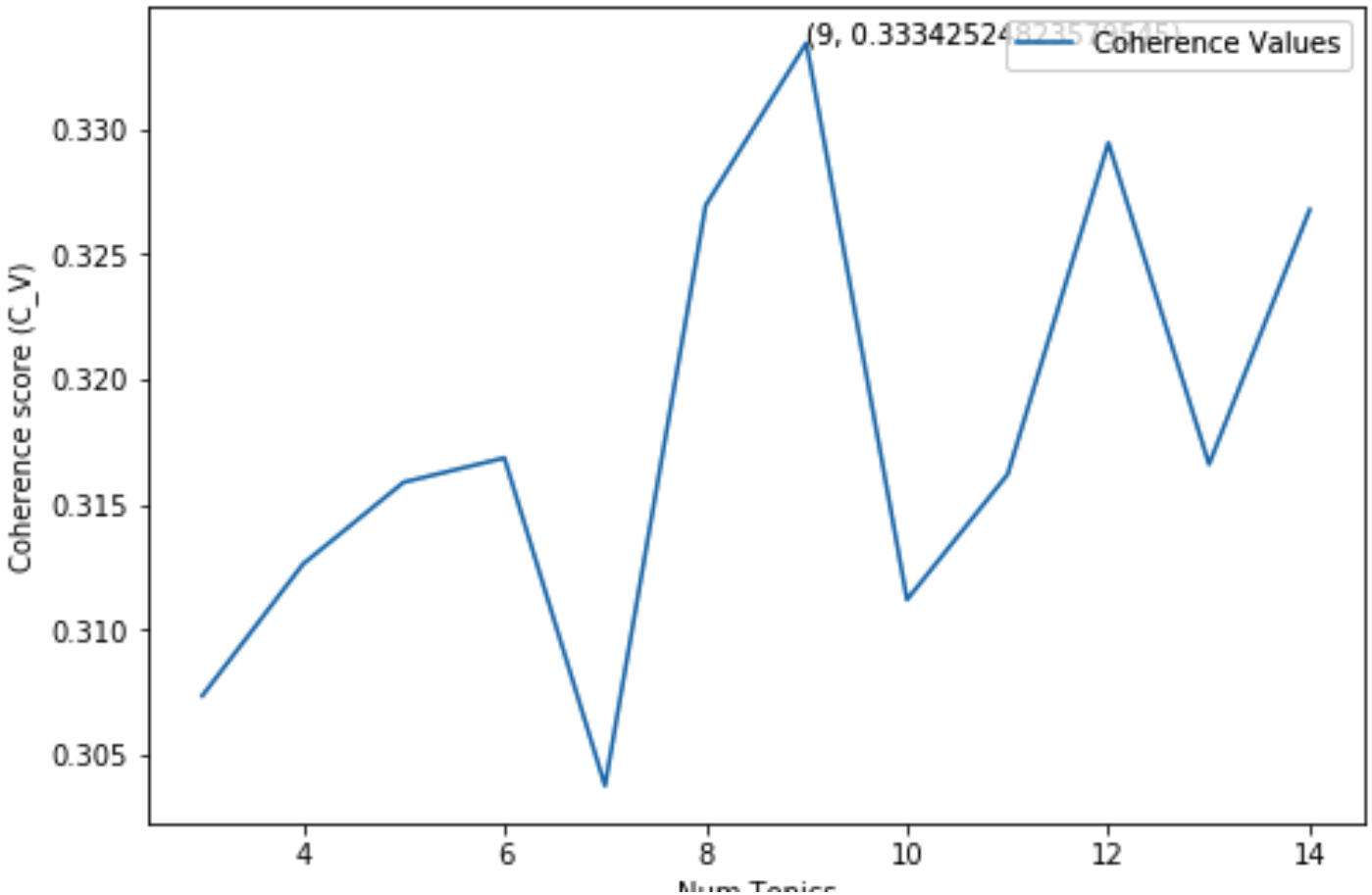
Topic Modeling Keywords per Year



Top Topic Modeling Keywords



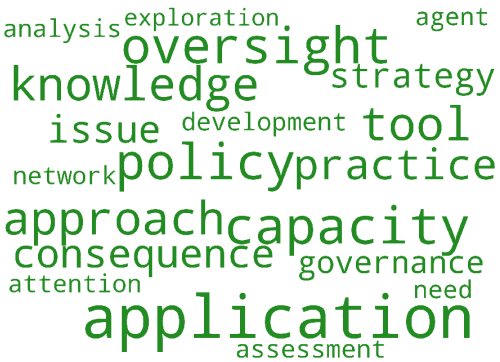
Why 9Topics?
Maximization of the CV coherence score



Wordclouds of Topics and Topic Modeling Keywords

Topic Modeling of the Web of Science bibliographic dataset on Anticipatory Governance

Topic 1 (Application Capacity)



Topic 2 (Policy Scenarios)

Topic 3 (Adaptation Policy)

Topic 4 (Innovation Governance)

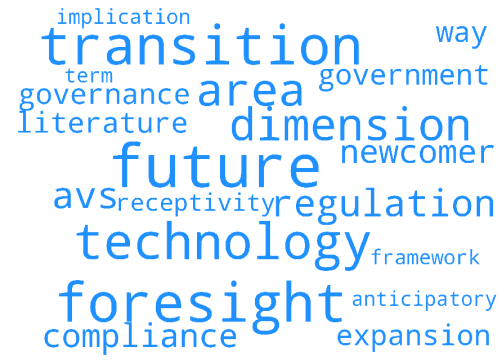
Topic 5 (Science Planning)

Topic 6 (Technology Governance)

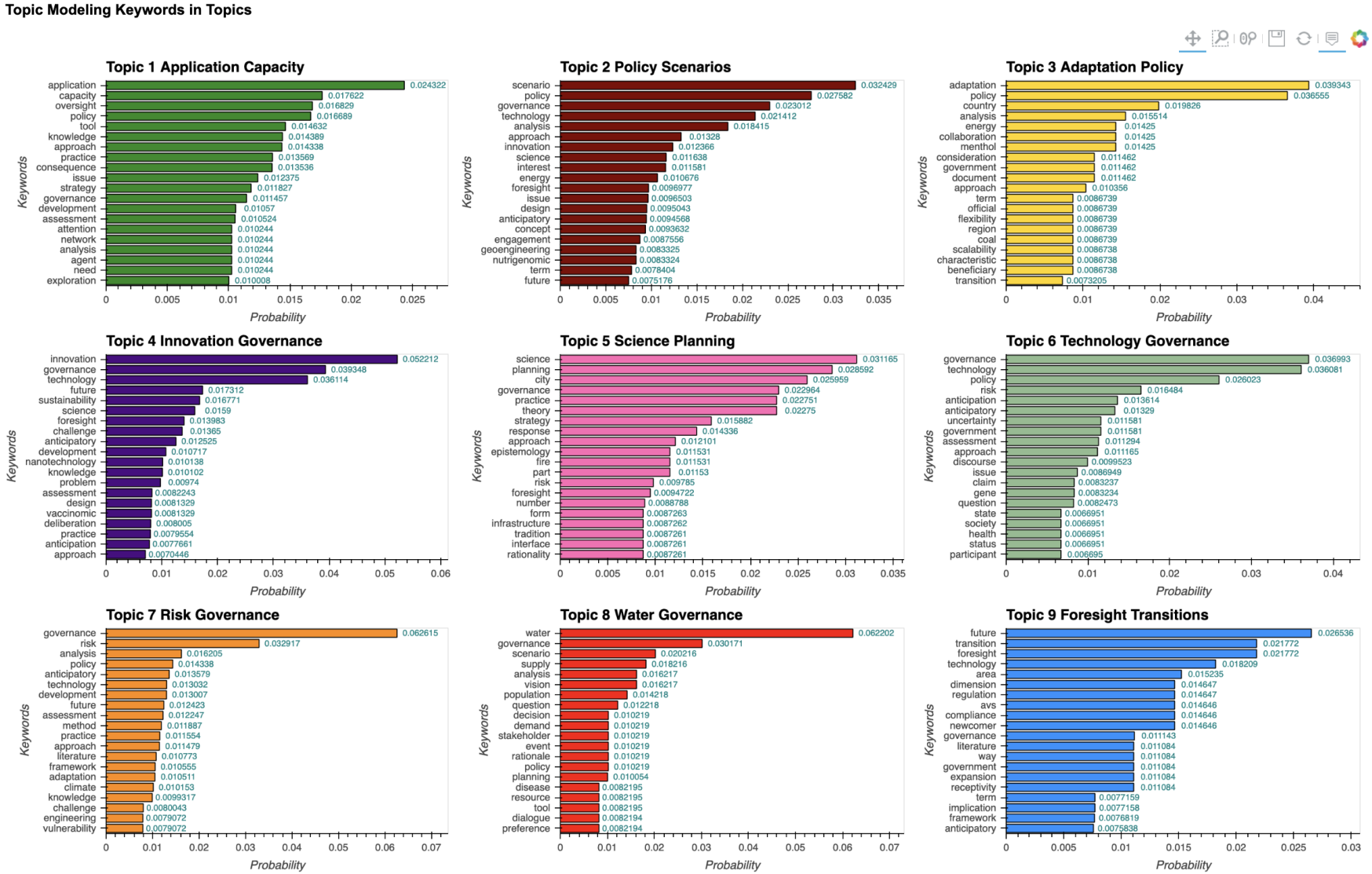
Topic 7 (Risk Governance)

Topic 8 (Water Governance)

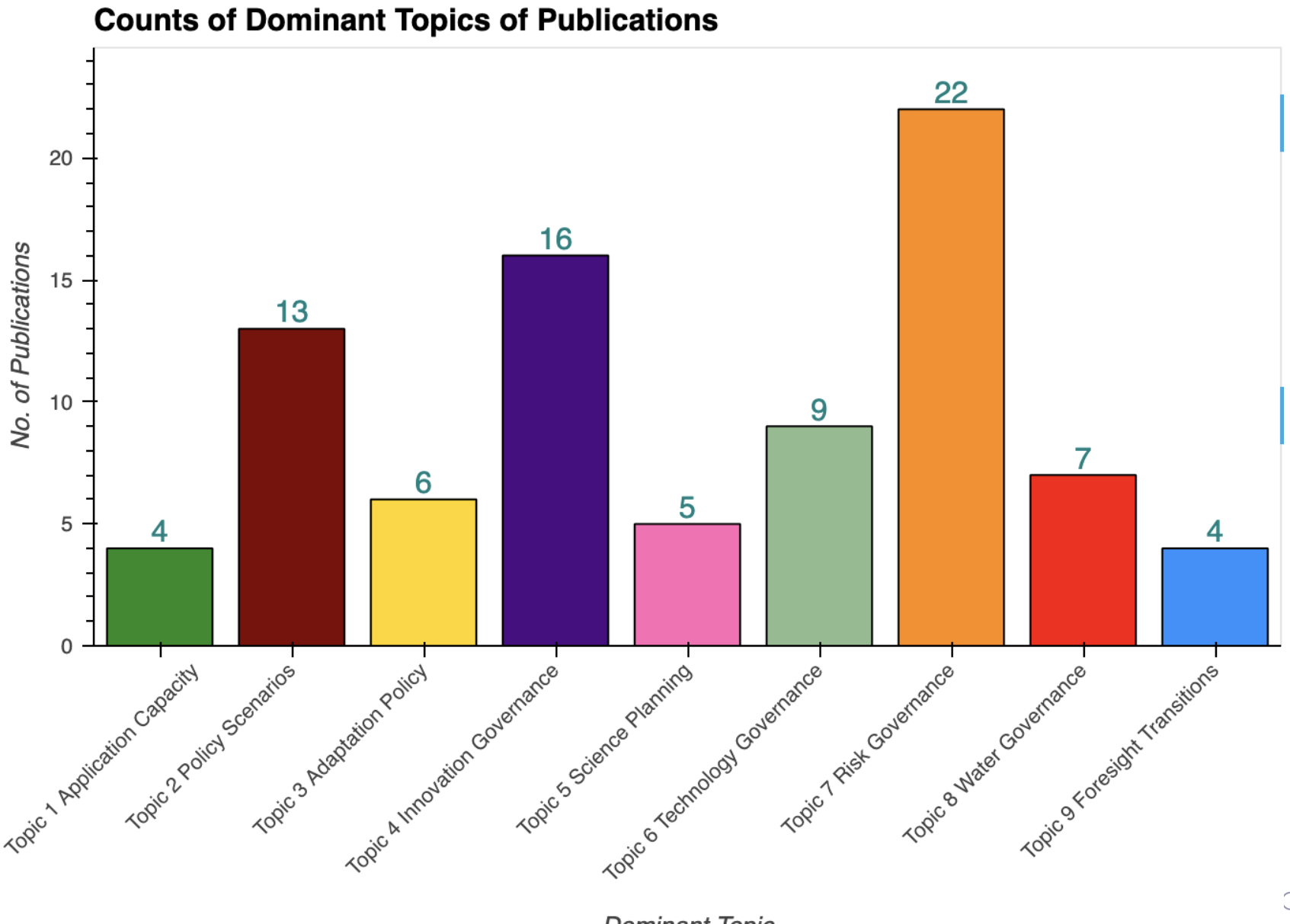
Topic 9 (Foresight Transitions)



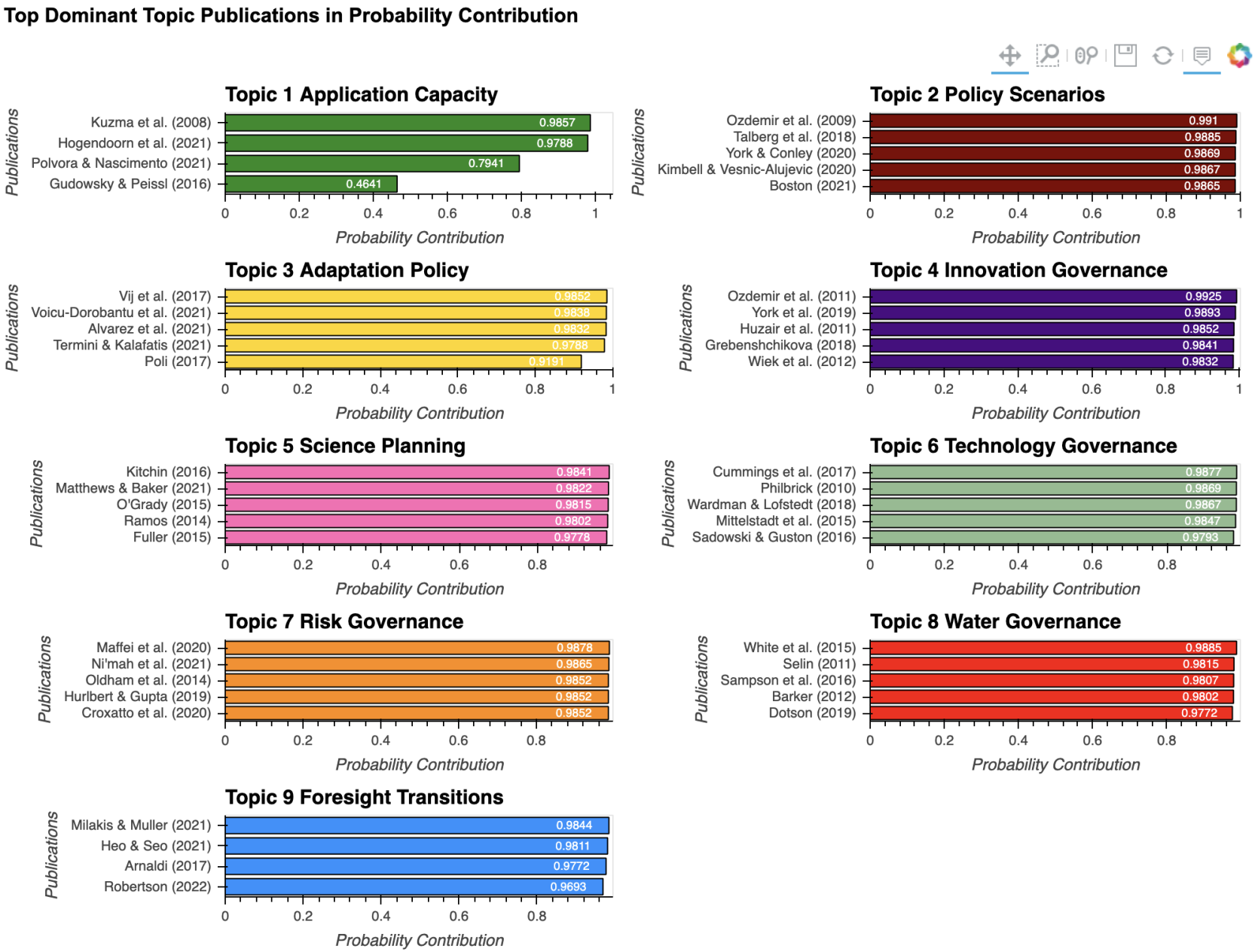
Weights of Topic Modeling Keywords



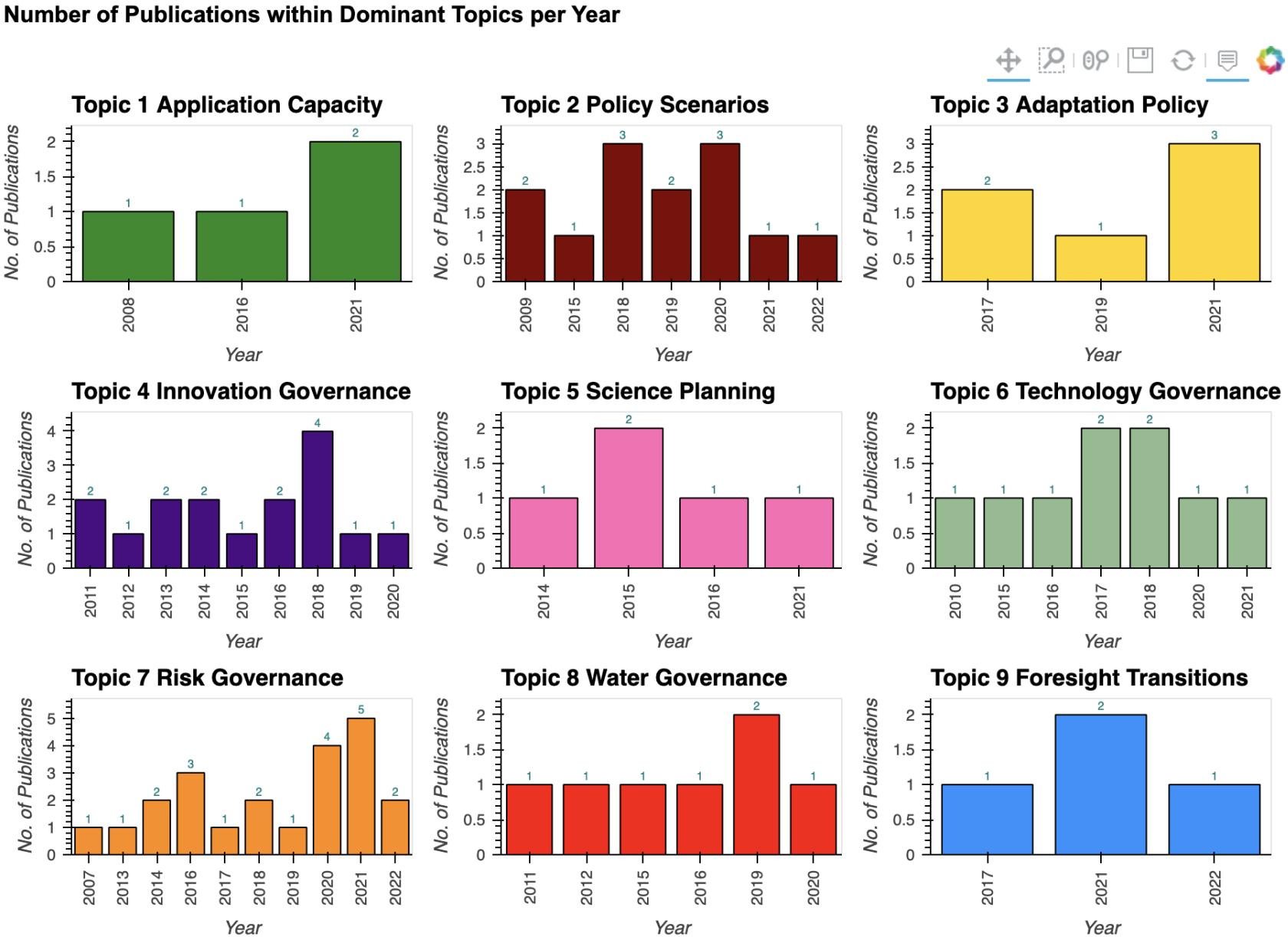
Dominant Topics of Publications



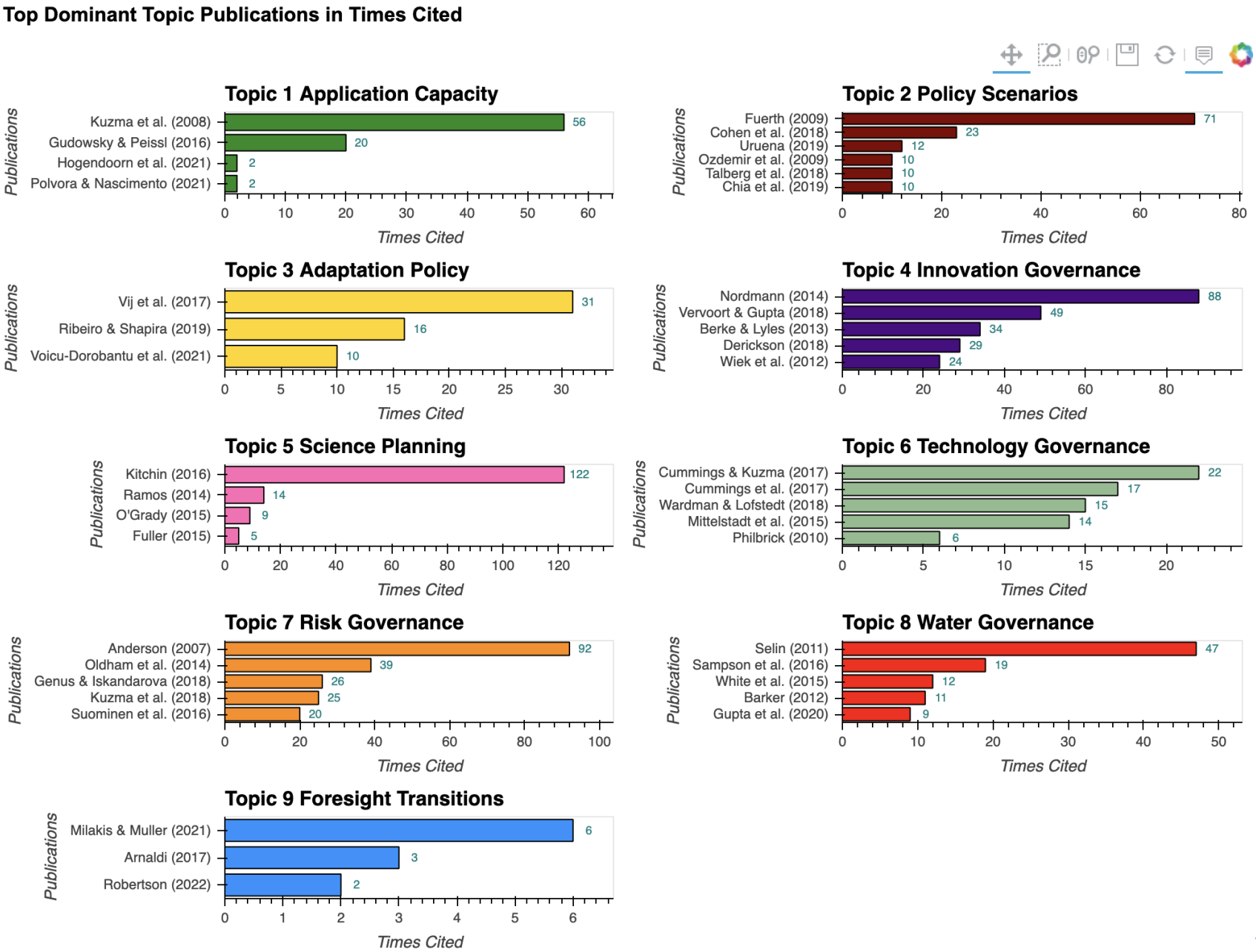
Top Dominant Topic Publications in Probability Contribution



Number of Publications within Dominant Topics per Year

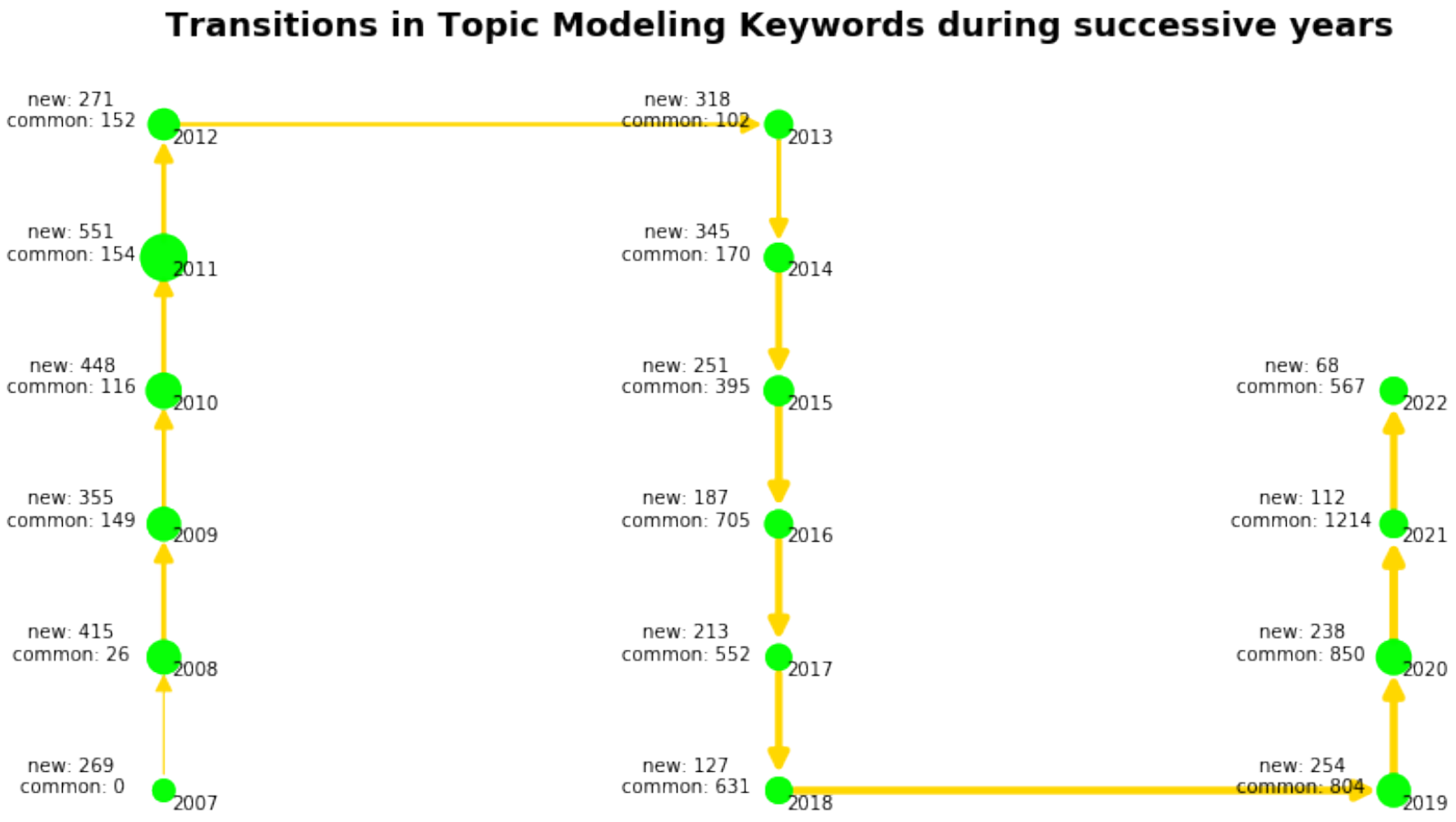


Top Dominant Topic Publications in Times Cited



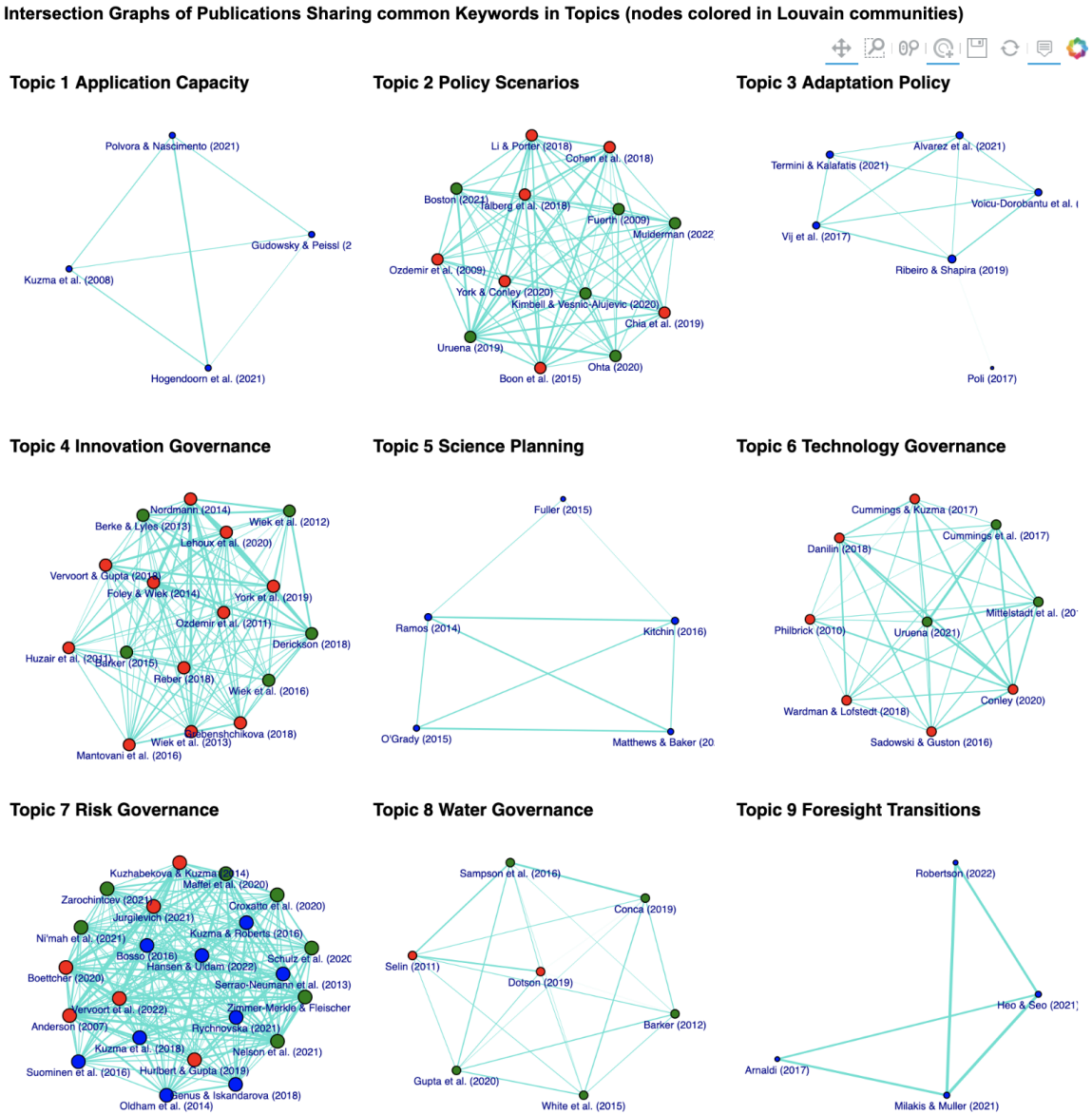
Transitions in Topic Modeling Keywords during successive years

new = new keywords in a year; common = common keywords with previous year



Link to an interactive graph of transitions

Intersection Graphs of Publications Sharing common Keywords in Topics



Link to the interactive intersection graphs