



Topic modelling tourism literature on innovation and technology

Presented by

Florian J Zach

Howard Feiertag Department of Hospitality and Tourism Management

Pamplin College of Business

Virginia Tech

Blacksburg, VA, USA



ENTER2020 International eTourism Conference

The 27th Annual Conference Organized by IFITT

Topic modelling tourism literature on innovation and technology

Florian J. Zach

Virginia Tech

Dejan Krizaj

University of
Primorska

Ajda Pretnar

University of
Ljubljana

Literature review

- Interplay of innovation and technology
 - how market stakeholders plan for and develop new technologies (Hargadon & Sutton, 1997),
 - adapt them (Chen & Tsou, 2007) and
 - manage the related innovation activities and outcomes (Salomo et al., 2008)
- Increase in both quantity and variety of research
 - (e.g. Navío-Marco et al., 2018; Trunfio & Campana, 2019)
- Recent reviews on innovation in tourism:
 - Focus on partnerships (Marasco, De Martino, Magnotti, & Morvillo, 2018)
 - Co-creation (Campos, Mendes, Valle, & Scott, 2018)
 - Focuses on cases (Carvalho & Costa, 2011)
 - AND: most often is limited to only “top” journals (from 5 to 25)

This study

- Purpose
 - Assessing the link between innovation and technology literature in tourism
 - Specifically to provide a first glimpse into the narrative of these studies
- Methodology
 - Include all studies on innovation and technology in a tourism context
 - Collect all publication abstracts
 - Apply natural language processing to abstracts
 - Analyze the processed data

Qualifying Publications

- keywords “innovation”, “tourism”, AND “technology”
- Title, abstract or keyword list
- English language
- Published before 31st December 2018.
- ScienceDirect database
- 318 abstracts from 1982 to 2018

Year	Count	Year	Count	Year	Count
1982	2	1999	2	2010	15
1983	1	2001	4	2011	18
1985	1	2002	3	2012	16
1988	1	2003	4	2013	17
1989	2	2004	3	2014	23
1993	1	2005	7	2015	24
1995	1	2006	4	2016	41
1996	1	2007	9	2017	37
1997	1	2008	12	2018	57
1998	1	2009	10		

Latent Dirichlet Allocation (LDA) to identify publication topics

- Among the most popular techniques of topic modelling
- Reverse-engineering the process of generating text
- **Key assumption:** each document is generated from a set of topics and further from a set of words defining these topics
- Setting a number of topics to each abstract and iteratively re-assigns words to topics based on their probability distributions in the document

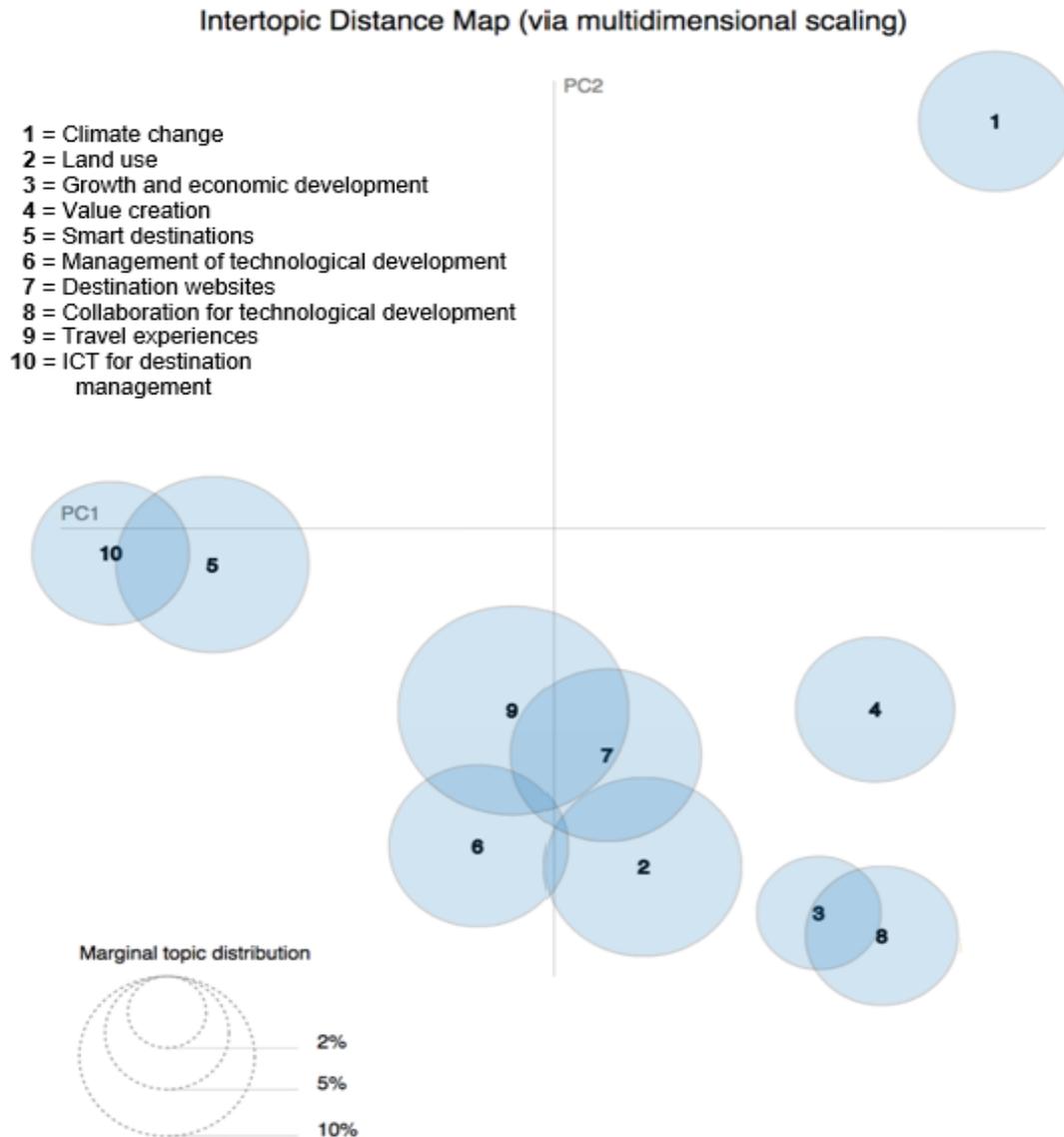
Preprocessing

- Lowercasing and tokenizing the words
- Lemmatized them with WordNet lemmatizer (Bird, Loper & Klein, 2009) >>> e.g. tourists to tourist
- Filtered out stopwords and corpus-specific words, which we manually defined
 - Mostly to the academic vocabulary, such as 'methodology', 'result', 'analysis' ...
- Final bag-of-words: kept the words occurring in more than 10% and less than 90% of documents.

Optimum number of topics

- Three methods for the number of topics between two and fifty:
 - Gensim's (Rehurek & Sojka 2010) log perplexity computes how well the topics fit new data; we used 70% of the data for training and 30% for testing
 - GridSearch algorithm from Python's sklearn suggested two topics
 - Gensim's Coherence Model (Roder, Both & Hinneburg, 2015) with Mallet LDA
- The results reported local maxima around 10 topics
> we ran LDA for 10 topics

10 topics



REVIEWER 1:

The authors seem to be very creative in giving their term groups **nice names**.

REVIEWER 2:

Apart from topic 1, which is very different, the other topics can't really be clearly distinguished on a first reading.

It requires **a manual interpretation** (and some ingenuity) to label them.

WE USED IT AS A PRE-SORTING TOOL, THEN TRIED TO INTERPRET EACH GROUP OF PAPERS MANUALLY.

5 clusters

A - Environment Effects

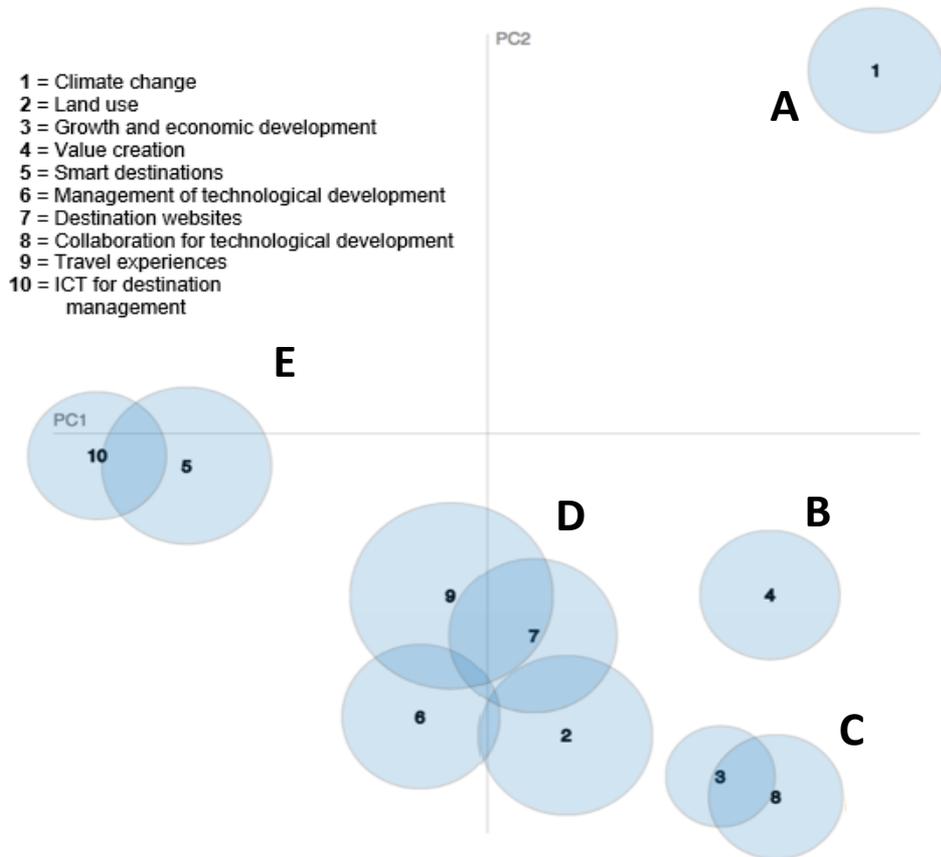
B - Value Creation

C - Business Operations

D - Destination Aspects

E - Tourism ICT

- 1 = Climate change
- 2 = Land use
- 3 = Growth and economic development
- 4 = Value creation
- 5 = Smart destinations
- 6 = Management of technological development
- 7 = Destination websites
- 8 = Collaboration for technological development
- 9 = Travel experiences
- 10 = ICT for destination management



Technology (E) generates value (B) with business operations (C) at destinations (D) leaving a footprint (A).

Conclusions

- Topic modelling on abstracts, despite the limitation of utilizing **only about up to 250 words**, is a fruitful exercise to gain a comprehensive **overview of an academic field**
- **Relatively low-effort activity** for new and established scholars to identify well covered research aspects and potential gaps
- The **investigated field was narrow** (innovation, tourism, technology) with only 318 abstracts, future research should develop in breadth and depth



ENTER2020 International eTourism Conference

The 27th Annual Conference Organized by IFITT

Topic modelling tourism literature on innovation and technology

Florian J. Zach

Virginia Tech

Dejan Krizaj

University of
Primorska

Ajda Pretnar

University of
Ljubljana