**The Power of Metadata**

**Structural Topic Modelling for Bibliometric Studies in Technology and Innovation Management**

Topical bibliometric studies within the Technology and Innovation Management research area

A close-up of a document

Description automatically generated with low confidence

Danie Ungerer

u29129398

**Abstract**

**Table of Contents**

[1 Introduction 1-1](#_Toc82366339)

[2 Methods and Data 2-1](#_Toc82366340)

[2.1 Data Pipeline 2-1](#_Toc82366341)

[2.2 Data Preparation 2-1](#_Toc82366342)

[2.3 Exploratory Data Analysis 2-1](#_Toc82366343)

[2.4 Bibliometric Analysis 2-1](#_Toc82366344)

[2.5 Structural Topic Modelling 2-1](#_Toc82366345)

[3 Results 3-3](#_Toc82366346)

[4 Discussion 4-10](#_Toc82366347)

[5 Conclusion 5-10](#_Toc82366348)

**List of Figures**

[Figure 2‑1: Diagnostic values by the number of topics 3-4](#_Toc82386720)

[Figure 2‑2: Comparing exclusivity and semantic coherence of chosen models 3-4](#_Toc82386721)

[Figure 2‑3: Chosen number of topics for labelling (Model K = 17) 3-5](#_Toc82386722)

[Figure 2‑4: Topic distribution in the corpus 3-5](#_Toc82386723)

[Figure 2‑5: Theta values per document 3-6](#_Toc82386724)

[Figure 3‑1: Topic proportions with top 10 words 4-8](#_Toc82386725)

[Figure 3‑2: Network Graph of Topic Correlation 4-1](#_Toc82386726)

[Figure 3‑3: Correlation plot between topics 4-2](#_Toc82386727)

[Figure 3‑4: Topic prevalence of time (normalised for incomplete 2021) 4-3](#_Toc82386728)

**List of Tables**

[Table 2‑1: Topic labelling comparison based on theta () values 2-4](#_Toc82383889)

[Table 3‑1: Topics identified and labelled 3-1](#_Toc82383890)

# Introduction

# Literature Review

# Methods and Data

## Data Pipeline

## Data Preparation

## Exploratory Data Analysis

## Bibliometric Analysis

## Structural Topic Modelling

### Finding Optimal Number of Topics

#### Performance Scores

* Held-Out Likelihood
* Residuals
* Semantic Coherence

Semantic coherence is a metric related to [pointwise mutual information](https://en.wikipedia.org/wiki/Pointwise_mutual_information); the core idea is that in semantically coherent models, the words that are most probable under a topic should co-occur within the same document. While there is a correlation between this metric and human judgement, models with fewer topics tend to have a high semantic coherence score.

* Exclusivity

#### Topic Number Diagnosis

Chart, line chart

Description automatically generated

Figure ‑: Diagnostic values by the number of topics

#### Choosing Optimal Topic Numbers

Chart, scatter chart

Description automatically generated

Figure ‑: Comparing exclusivity and semantic coherence of chosen models

#### Model with K = 17 topics

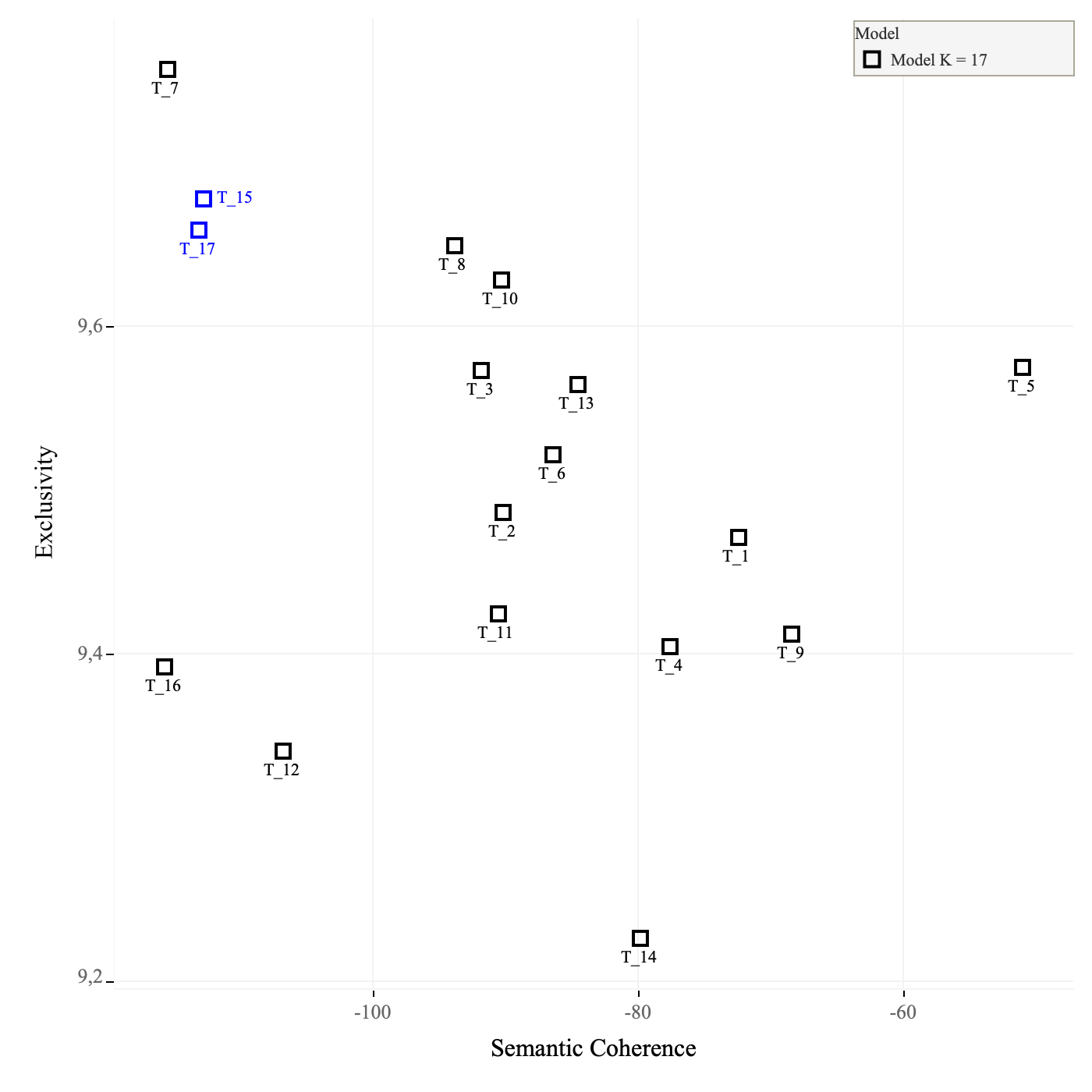


Figure ‑: Chosen number of topics for labelling (Model K = 17)

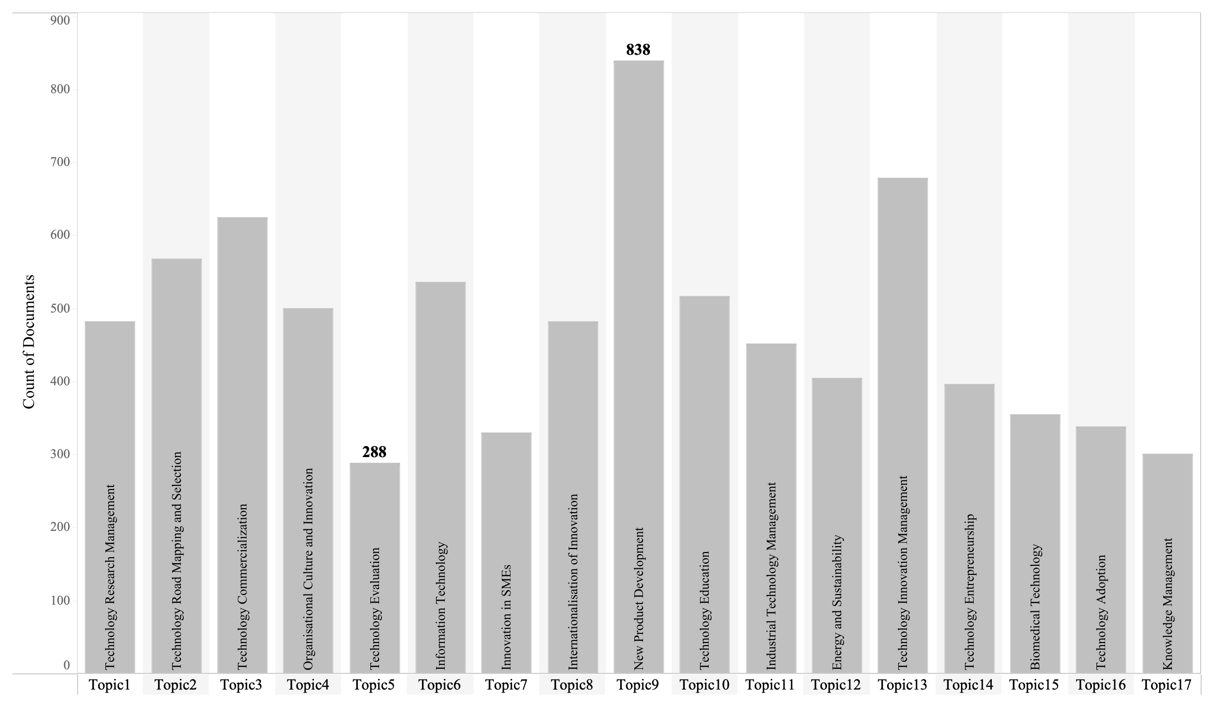


Figure ‑: Topic distribution in the corpus

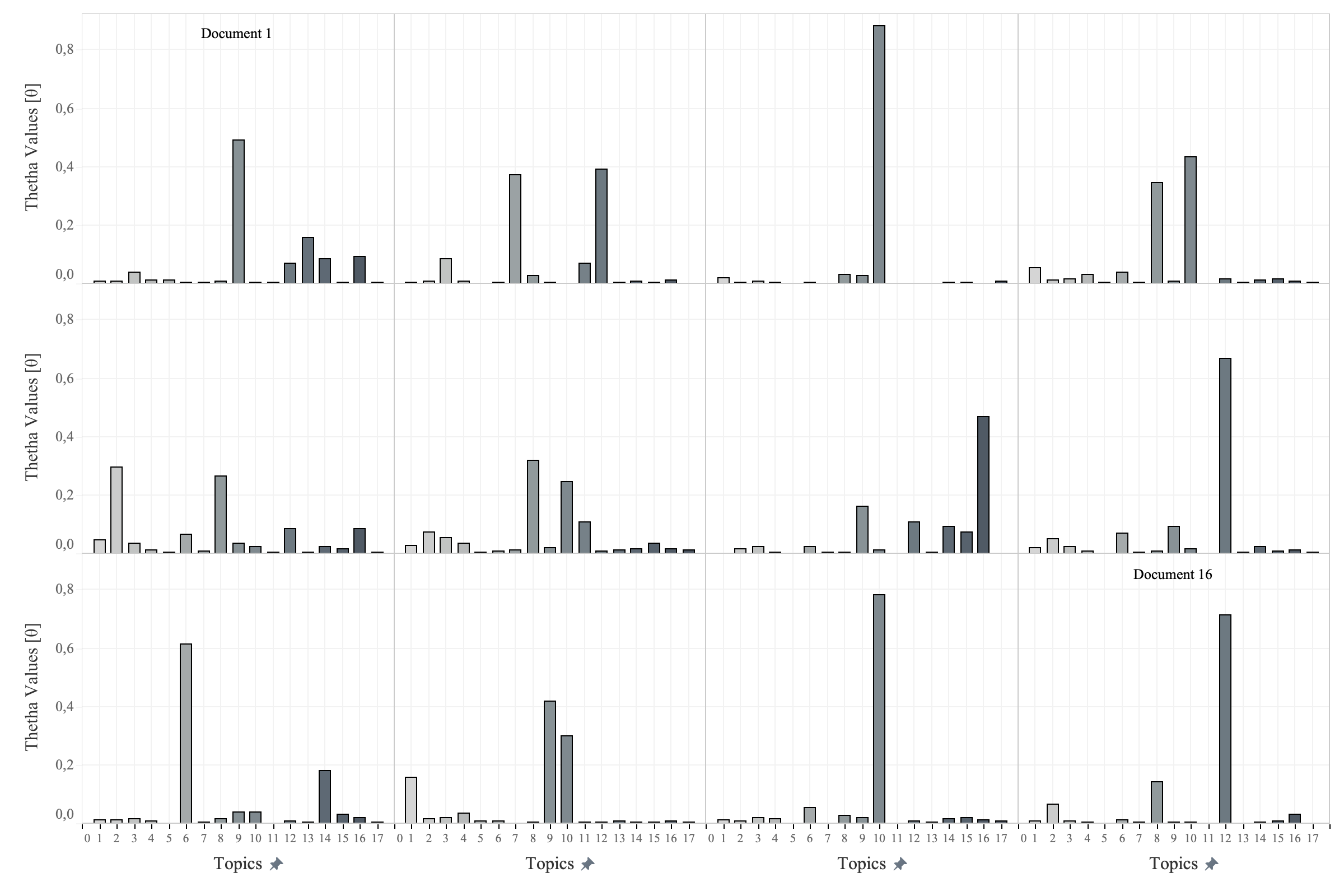


Figure ‑: Theta values per document

Table ‑: Topic labelling comparison based on theta () values

| Document | Title | Abstract | Labelled Topic |
| --- | --- | --- | --- |
| 4 | A PILOT SURVEY OF COURSES IN TECHNOLOGY TRANSFER OFFERED AT SELECTED COLLEGES AND UNIVERSITIES IN THE UNITED STATES | THIS SURVEY WAS DEVELOPED TO DETERMINE THE NATURE OF THE PROGRAMS AND COURSES IN TECHNOLOGY TRANSFER NOW IN EXISTENCE IN THE UNITED STATES. A PILOT STUDY APPROACH WAS USED, WITHOUT OPERATIONALIZING A DEFINITION, TO AVOID EXCLUDING POSSIBLE TECHNOLOGY TRANSFER ACTIVITIES. RESULTS INDICATE TWENTY COLLEGES AND UNIVERSITIES OFFERING COURSES IN A WIDE RANGE OF ACADEMIC SETTINGS, INCLUDING ENGINEERING, BUSINESS, PSYCHOLOGY, SOCIOLOGY, POLITICAL SCIENCE, ENVIRONMENTAL STUDIES AND AGRICULTURE. WITH THE MAJORITY OF RESPONDENTS, TECHNOLOGY TRANSFER IS STUDIED AS A COMBINATION OF THE HISTORY OF SCIENCE AND TECHNOLOGY, MANAGEMENT, INFORMATION SCIENCE, AND COMMUNICATION. IN THIS SENSE, IT APPEARS TO BE AN INTERDISCIPLINARY STUDY WITH A STRONG DIRECTION TOWARD COMMUNICATION PROTOCOLS AND MANAGEMENT SKILLS AS PRIMARY OUTCOMES. 1978 TECHNOLOGY TRANSFER SOCIETY. | Topic10 |
| 15 | ESTABLISHING A TRAINING PROGRAM FOR THE INTERNAL AUDIT DEPARTMENT | CONSTANTLY INCREASING DEVELOPMENTS IN BUSINESS, INCLUDING COMPUTER TECHNOLOGY, MANAGEMENT DECISION AIDS AND THE PROLIFERATION OF FINANCIAL REPORTING REQUIREMENTS HAVE MADE THE NEED FOR CONTINUING EDUCATION PROGRAMS INCREASINGLY IMPORTANT TO THE ENTIRE FINANCIAL COMMUNITY AND INTERNAL AUDIT DEPARTMENTS. A CONTINUING EDUCATION PROGRAM WILL PROVIDE INTERNAL AUDITORS WITH CURRENT KNOWLEDGE TO EFFECTIVELY ACCOMPLISH THEIR MISSION OF SERVING MANAGEMENT. IN ORDER TO DEVELOP AN EFFECTIVE TRAINING PROGRAM FOR THE INTERNAL AUDIT STAFF, THE TRAINING PROGRAM GOALS SHOULD BE FORMALLY STATED AND PRESENTED TO THE INTERNAL AUDIT STAFF. THE GOALS SHOULD BE STATED IN SUCH A MANNER THAT STAFF MEMBERS WILL ADOPT THEM AS THEIR OWN. EXAMPLES OF GOALS ARE DISCUSSED. THE FUNDS AND TIME DEVOTED TO INTERNAL AUDIT TRAINING WILL YIELD A HIGH RETURN TO MANAGEMENT IN THE FORM OF A COMPETENT INTERNAL AUDIT STAFF WHICH PRODUCES EFFECTIVE RECOMMENDATIONS AND A STORE OF TALENT THAT CAN BE CALLED UPON FOR SPECIAL ASSIGNMENTS AND ADDED RESPONSIBILITIES. | Topic10 |

# Results

## Identified Topics and Intuative Meanings

Text, letter

Description automatically generated

Figure ‑: Topic proportions with top 10 words

Table ‑: Topics identified and labelled (follow hyperlinks in No. column for detailed wordcloud and abstract summation)

| No. | Topic Label | Top-10 Words | FREX | *Lift* | Score |
| --- | --- | --- | --- | --- | --- |
| [1](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_1) | Technology Research Management | research, management, literature, review, field, studies, analysis, study, article, future | articles, papers, journal, literature review, review, bibliometric, journals, | intellectual structure, jpim, management journals, papers published, web science, bibliometric, bibliometric analysis, co-citation | literature, articles, journals, papers, review, journal, bibliometric, literature review |
| [2](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_2) | Technology Road Mapping and Selection | technology, technology management, management, decision, technologies, analysis, technological, model, method, patent | road mapping, decision making, ahp, printing, technology road mapping, patent, decision, roadmap, roadmaps, 3d printing | 3d printing, road mapping, technology intelligence, technology road mapping, business decision, citation networks | patent, technology, decision, patents, technology management, decision making, technology road mapping, patent analysis, roadmaps |
| [3](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_3) | Technology Commercialization | business, market, companies, industry, strategy, change, technology, strategic, competitive, digital | disruptive, business models, business model, markets, technological change, competition, disruptive innovation, change, competing, business strategy | corporate accelerators, corporate applications, enter market, future thinking, iamot rights, implementing digital | business, disruptive, business model, digital, market, business models, companies, competitive, markets, newspace |
| [4](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_4) | Organisational Culture and Innovation | study, management, factors, results, performance, research, data, organizational, success, survey | tqm, questionnaire, satisfaction, employee, respondents, quality management, project success, success factors, variables, data collected | identified factors, intrinsic motivation, quality dimensions, random sampling, team performance, csfs, erp implementation, job satisfaction, oim, project success | tqm, structural equation, employees, satisfaction, factors, leadership, questionnaire, employee, erp, organizational |
| [5](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_5) | Technology Evaluation | design, purpose, study, findings, approach, research, implications, limited, methodology | design methodology, methodology approach, research limitations, originality paper, npd, practical implications | lsps, cocreation, innovation emerald, lego, methodology approach, research limitations, atypical package | npd, design methodology, originality, methodology approach, research limitations, practical implications |
| [6](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_6) | Information Technology | information, data, management, system, systems, technology, information technology, service, network | cloud, big data, ai, cloud computing, information security, computing, security, artificial intelligence, information system, artificial | ai, cloud, data warehouse, mpls, cloud computing, cobit, data mining, query, ai audit, ai-based | cloud, big data, information, security, information technology, cloud computing, computing, information security, ai, information systems |
| [7](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_7) | Innovation in SMEs | enterprises, manufacturing, chain, supply, smes, industry, small, management, supply chain | supply chain, small medium, enterprises smes, chain management, medium-sized enterprises, smes, small medium-sized, chain, medium enterprises, amt | chain management, chain performance, enterprises smes, small medium, small medium-sized, advanced manufacturing, agricultural investment, amt, amts | smes, supply chain, supply, enterprises, chain, korean, manufacturing |
| [8](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_8) | Internationalisation of Innovation | technology, research, policy, public, countries, science, government, development, national, transfer | technology transfer, property, intellectual property, government, private, licensing, public sector, ip, science business | public research, access licensing, advanced countries, advanced research, collaborator, commercialization university, darpa | technology transfer, policy, government, ip, public, intellectual property, science technology, property, universities |
| [9](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_9) | New Product Development | process, product, development, project, management, projects, design, paper, approach, processes | lean, design process, project management, phases, development process, business process, development projects, software development, pss, front end | architecting, fuzzy front, pd, plm, process automation, product concepts, software process, uncertainty ambiguity, agile innovation, agile project | product, project, product development, projects, project management, process, lean, engineering, pss, manufacturing |
| [10](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_10) | Technology Education | technology, management, engineering, education, students, program, technology management, learning, programs, university | courses, graduates, teachers, faculty, student, graduate, engineering technology, engineering education, management education, education | academic programs, active learning, agstem, alumni, atmae, btm, built environment, capstone, careers, classrooms | students, education, engineering, courses, program, student, teaching, programs, educational, curriculum |
| [11](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_11) | Industrial Technology Management | firms, performance, firm, capabilities, capability, product, technology, management, study, relationship | dynamic capabilities, firm performance, family firms, absorptive, management capability, absorptive capacity | moderating role, technology sourcing, absorptive capability, absorptive capacities, adaptive capability, affect firm, affect firm's, capabilities firm, capabilities theory | firms, innovation performance, dynamic capabilities, firm's, capability, firm performance, family firms, product, absorptive |
| [12](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_12) | Energy and Sustainability | management, energy, construction, technology, environmental, system, development, industry, production, control | energy, water, renewable, renewable energy, emissions, tech publications, trans tech, construction, | acid, acrylic, antenna, arrestor, biogas, cereals, coal, coal mine, coastal, combustion | energy, construction, water, renewable, renewable energy, emissions, co2, gas, fuels, solar |
| [13](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_13) | Technology Innovation Management | innovation, innovation management, management, open, process, paper, open innovation, innovations, innovation process, companies | open innovation, radical innovation, innovation practices, service innovation, innovation process, model innovation, total innovation, radical, innovation processes, innovation | characteristics innovation, concept open, cops innovation, ff, innovation laboratories, innovation open, radical innovation, ri, discontinuous innovation, innovation companies | innovation, open innovation, innovation management, innovation process, open, innovation processes, product innovation, oi, model innovation, innovation activities |
| [14](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_14) | Technology Entrepreneurship | development, innovative, management, innovation, economic, research, main, analysis, economy, article | innovative development, economa | economic entities, urban planning, association ibima, de la | innovative development, innovation, sub‑regional, innovative activity |
| [15](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_15) | Biomedical Technology | health, technology, medical, healthcare, management, care, clinical, equipment, maintenance | health, medical, healthcare, care, clinical, hospitals, health care, patient | health information, health technology, aami, care delivery, care technology, caregivers, ced, clinical engineering, clinical engineers | health, medical, clinical, healthcare, clinical engineering, care, hospitals, patient, health care, hospital |
| [16](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_16) | Technology Adoption | adoption, technology, time, cost, case, model, mobile, service, internet, services | technology adoption, technology acceptance, outsourcing, acceptance, farmers, teaching notes, retail, yield, rfid, tam | familiarity, synopsis case, winter's tale, acceptance model, atma, code css, contact library, court, details email, educators contact | mobile, rfid, consumers, teaching notes, retail, technology acceptance, winter's, tam, outsourcing, farmers |
| [17](https://github.com/danievanmopanie/mit807_power_of_metadata/tree/main/artwork/topic_17) | Knowledge Management | knowledge, management, networks, network, organizational, collaboration, knowledge management, learning, collaborative, paper | knowledge management, knowledge sharing, knowledge transfer, knowledge creation, knowledge, innovation networks, knowledge innovation, km, information knowledge, organizational learning | collaborative networks, ki, kibs, knowledge creation, knowledge workers, management km, sharing knowledge, absorbed, acap, bda | knowledge, knowledge management, networks, km, knowledge sharing, ict, knowledge innovation, knowledge creation, network, collaborative |

Chart

Description automatically generated

Figure ‑: Network Graph of Topic Correlation

Chart

Description automatically generated

Figure ‑: Correlation plot between topics

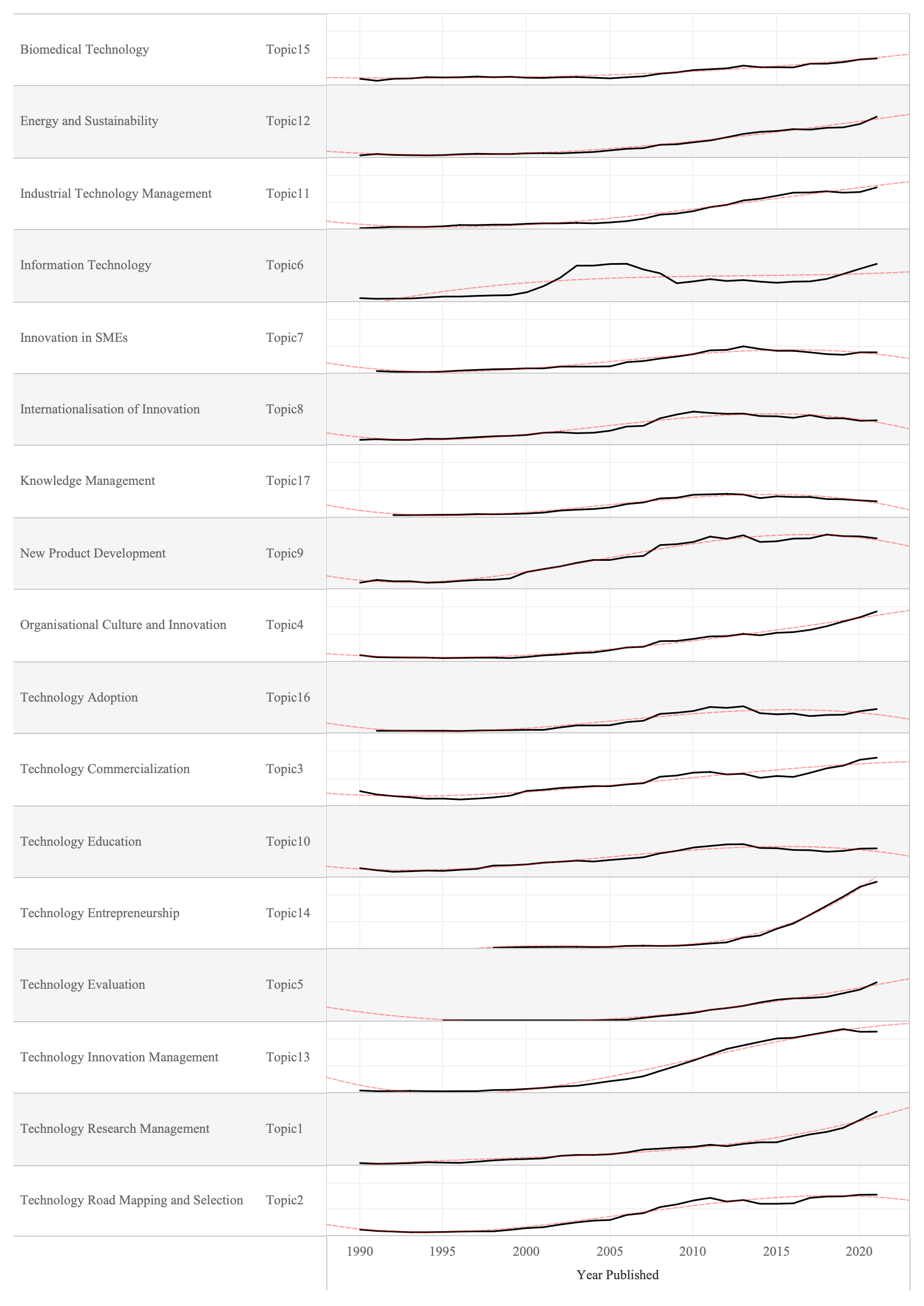


Figure ‑: Topic prevalence of time (normalised for incomplete 2021)

# Discussion

# Conclusion

# References