Asian Journal of Technology Innovation

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rajt20

News big data analysis of international start-up innovation discourses through topic modelling and network analysis: comparing East Asia and North America

Kyeo Re Lee, Jang Hyun Kim, Jaeyeon Jang, Jeewoo Yoon, Dongyan Nan, Yonghwan Kim & Byungjun Kim

To cite this article: Kyeo Re Lee, Jang Hyun Kim, Jaeyeon Jang, Jeewoo Yoon, Dongyan Nan, Yonghwan Kim & Byungjun Kim (2022): News big data analysis of international start-up innovation discourses through topic modelling and network analysis: comparing East Asia and North America, Asian Journal of Technology Innovation, DOI: 10.1080/19761597.2022.2134154

To link to this article: https://doi.org/10.1080/19761597.2022.2134154

Published online: 19 Oct 2022.

ASIAN JOURNAL OF TECHNOLOGY INNOVATION
https://doi.org/10.1080/19761597.2022.2134154

RESEARCH ARTICLE

News big data analysis of international start-up innovation discourses through topic modelling and network analysis: comparing East Asia and North America

Kyeo Re Lee [a], Jang Hyun Kim [a,b,c], Jaeyeon Jang [a], Jeewoo Yoon [c], Dongyan Nan [a,b,d], Yonghwan Kim [e] and Byungjun Kim [f]

[a]Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, Korea; [b]Department of Interaction Science, Sungkyunkwan University, Seoul, Korea; [c]Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, Korea; [d]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore; [e]NAVER Green Factory, Seongnam, Korea; [f]Center for Digital Humanities and Computational Social Sciences, Korea Advanced Institute of Science and Technology, Daejeon, Korea

ABSTRACT

It is important to examine how start-up innovation is emerging in the information society. The media is a window through which innovation is highlighted in which fields. We conducted a comparative study with North America to comparatively analyze how startup innovation in East Asia appeared. This study used a set of computer analysis methods including Dirichlet-multinomial regression Topic Model and Topic network analysis. News articles from 2000 to 2019 were collected from East Asia and North America and were analysed on the topic of start-ups. The results indicated that the discourse of start-ups from East Asia and North America in the 2000s and 2010s showed distinctly different trends. The East Asia media changed its focus from an innovation economy to an emerging industry: mobility & energy, while the North American media showed a change from revenue sources to the benefit of innovation. In the discourse of start-ups in East Asia, the government-centred innovative economy was emphasised in the 2000s, and expectations for emerging industries such as mobility and energy increased after the 2010s. Meanwhile, in the discourse of start-ups in North America, revenue sources were emphasised in the 2000s, and the proportion of beneficiaries of innovation increased after the 2010s.

KEYWORDS

News discourse analysis; start-up innovation; text mining; network analysis; topic modelling

1. Introduction

Companies have always changed and developed in accordance with the times, and the discovery and utilisation of new technologies has had a great impact on the development of companies. It is possible to examine how changes in entrepreneurship and social interests have been changed and emphasised through how news reports by media have changed.

CONTACT Byungjun Kim kuntakim88@gmail.com Center for Digital Humanities and Computational Social Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea

Technological innovation may appear differently depending on cultural differences. To examine the changes in innovation, news articles were collected from various media around the world, focusing on the period when Internet-based businesses evolved and the use of smartphones in the corporate ecosystem began spreading. Articles from 2000 to 2019 were collected and divided between the North American region, which has traditionally excelled in the development and utilisation of the Internet technology, and the emerging East Asian region. There are two major sectors to make regional innovation systems (RIS) (Yoon et al., 2015). The two divisions are divided into Mature entrepreneurial RIS and Still-evolving entrepreneurial RIS. Geographically, Mature entrepreneurial RIS is represented by the United States, and Still-evolving entrepreneurial RIS refers to the East Asian region (Yoon et al., 2015). This study also compares the media reports in North America and East Asia. Comparing the perspectives of the two cultures helps track the effects of culture on the perception of start-ups. The scale of start-up investments is increasing in both North America and China, as seen in Figure 1.

From this graph, it can be seen that the gap between cultures in terms of start-up investment is gradually narrowing over time. To examine start-up trends and RIS, it is necessary to review a wealth of data beyond simple indicators; however, news provides refined data. Furthermore, the use of international news data makes it possible to know the global trend of start-ups and RIS.

The rest of this paper is organised as follows. The background section provides the theoretical and methodological background for this research. It includes past literature on start-ups, agenda setting theory, and computational approach to agenda setting, and research questions. The methods section explicates the Dirichlet-multinomial regression Topic Model and generalised Dirichlet-multinomial regression Topic Model used in this study. The results section reports the findings of this study. The discussion and suggestions for future research section interprets the results and provides suggestions for research that deepen the findings of this research.

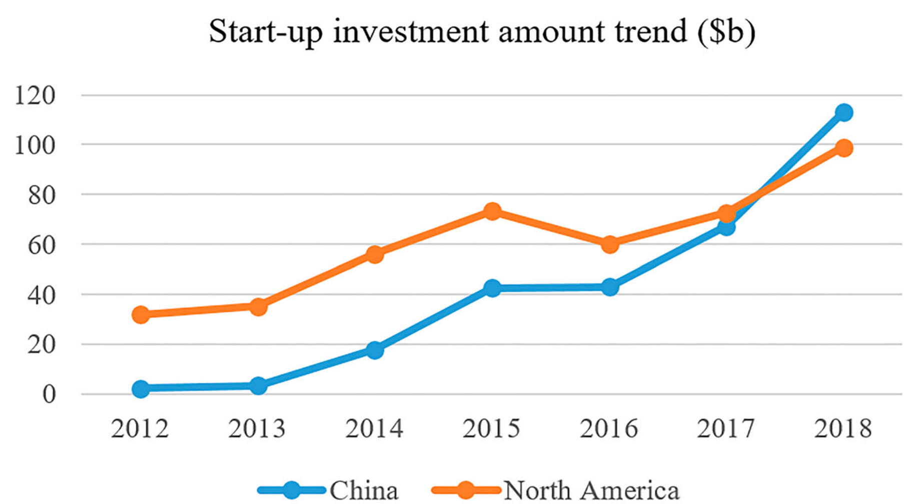


Figure 1. Start-up investment trend (Source: CB Insights quoted in IIT Trade Focus).

2. Background

2.1 Start-ups at the edge of innovation

A start-up's role in the industry is frequently compared to an engine of innovation in the field of management. The conventional definition of a start-up refers to Steve Blank's principles (2010), where he nomenclatures that start-ups are companies, partnerships, or temporary organisations repeatedly aiming at a 'scalable business model'. Start-ups are companies setup by entrepreneurs, usually with less capital, fewer human resources, less legal expertise, and less strategic cooperation than established companies. However, they are the heart of innovation because mobility of resources and alignment of incentives are crucial drivers for market growth breakthrough (Freeman & Engel, 2007). Colombelli et al. (2020) conceptualised the term 'innovative start-up' as a start-up company with a high degree of R&D, a high quality of human capital, or intellectual property rights. Furthermore, several scholars have tried to explore how to enhance start-up innovation and corporate performance (e.g. Huang et al., 2012; Guckenbiehl et al., 2021). Due to the above conceptual definition, startup innovation encompasses a variety of innovative activities based on the flexibility of startups, and through this, it is meaningful in achieving the purpose that the company seeks.

As of 2020, digital transformation has created more innovation opportunities for firms, as digital technologies encourage them to upgrade their processes, goods, and services into the new paradigm of competitiveness (Hinings et al., 2018). With the advent of the industry 4.0 paradigm, start-ups have become the fundamental source of knowledge in the society, and despite being unsuccessful, their knowledge can be valuable for other foundations or persons (Spender et al., 2017).

Open innovation (OI) is a widely used term in academy of social science, which means that society should utilise both external and internal idea/resources for its growth (Chesbrough, 2003). This idea seems an extended version of Public-Private Partnership (PPP) frameworks, often called as Public-Private-People Partnership (4P) by managerial studies (Majamaa, 2008). Entrepreneurs tend to act together (Bogers et al., 2018); their entrepreneurial activity may affect many facets of society; thus, entrepreneurship is a key contributor to the spread of innovation culture. In this context, 'practical and managerial backup to recently developed tech-based firms, spin-offs and start-ups (Jugend et al., 2020)' is, among recent innovation studies, one of the most researched subjects. Past studies have showed that governmental support has had a positive impact on the development of small firms (Jugend et al., 2020).

2.2 Innovation factors and entrepreneurial regional innovation system (RIS)

Corporate innovation refers to the introduction of new ideas or work methods to pursue the goals set by the company. And these innovations can be applied quickly in small businesses such as start-ups. There are a number of factors contributing to this innovation. Innovation factors include R&D (Avermaete et al., 2004), Cooperation (Kim, 2005), absorption capacity (Cohen & Levinthal, 1990), Appropriability (Teece, 1986), government support (Kang, 2019), regulation (Ahn et al., 2018) and others. Start-ups innovation is differentiated by regional innovation systems (RIS) (Cooke & Leydesdorff, 2006). The RIS refers to the system in which innovation and economic development occur in small geographical entities of a country based on the control of regional governments (Iammarino, 2005; Cooke et al., 1997; Lew & Park, 2021). Lau and Lo (2015) found that three factors of RIS (i.e. value chain information source, regional innovation initiative, knowledge-intensive business service) have positive relationships with enterprises' innovation performance. Lew and Park (2021) also indicated that internalisation of RIS can lead to better economic development and sustainable development.

Boston Consulting Group (BCG) selected the 50 most innovative companies in the world for 2020 and analyzed their success factors (BCG, 2020). Many of the selected companies were from North America or East Asia. There are 10 innovation factors derived from the analysis of more than 1000 companies in BCG. The innovation elements are as follows: Innovation Ambition, Innovation Domains, Innovation, Governance, Performance Management, Organisation and Ecosystems, Talent and Culture, Idea-to-Market Fit, Project Management, Funnel Management, Portfolio Management (BCG, 2020).

Mature entrepreneurial RIS appears in North America, and start-ups are naturally created by individual actors (Cooke & Leydesdorff, 2006). On the other hand, in the case of Still-evolving entrepreneurial RIS, it appears in East Asia, and start-ups have the characteristic of being created by the government (Yoon et al., 2015). The Table 1 below summarises these characteristics.

This study aims to prove that the trends in media reports on start-ups over time would help gain insights into innovation by adopting media research methods in the field of communication, along with big data analysis. Especially, from the perspective of RIS, this study aims to examine how the discourse on start-ups in North America and East Asia is structured.

2.2.1 Computational approach to agenda-setting

Research on the concept of framing has been conducted in many fields of social science such as journalism, psychology, and sociology. Cohen stated that people perceive what the news emphasises as important and argued that the media emphasised not 'what to think', but 'what to think about' for people (Cohen, 1963). Framing theory posits that the selective suggestion of information can affect viewpoints, behaviours, hopes, and judgements (Chong & Druckman, 2007). Miller (1997) and Kim et al. (2007) argued that it is better to analyse media frames using algorithms rather than traditional human coding and insisted that it is possible to grasp the media frame based on the patterns of keyword map retrieved by a computer. Online news promotes an interaction with news audiences or creates an environment where citizens can exchange opinions more actively (Nah et al., 2020). Feezell (2018) argued that the traditional method of analysing

Table 1. Entrepreneurial RIS (Compiled from Yoon et al., 2015).

Category	Mature entrepreneurial RIS	Still-evolving entrepreneurial RIS
Geographic region	U.S. and other Anglo-American economies	East Asia
Origin of start-ups	Spontaneous-driven	Government-led
Key role of start-ups	Small business entrepreneurship, scalable start-up entrepreneurship	Promoting huge corporations' entrepreneurial efforts
Key factors of start-ups	Individual factors (entrepreneurs, venture capitalists, researchers, incubators)	Government (space, government research institutes)

agenda setting needs to be changed in the digitalisation era, and the study demonstrated that exposure on an issue by the SNS channel increases its perceived salience.

Probabilistic topic models are techniques used to reveal hidden thematic shapes in text by protruding each paper into a low dimensional space bridged by a set of topics (Nguyen 2015). Son and Hong (2019) argued that the trend of issues can be grasped by analysing the tendencies of various topics through the conflict frame study on regional issues.

2.3 Data-driven approach in media trend research and forecasting

Comparing different countries' media is an emerging area of study in many disciplines including political science, communication, and management. Kim and Barnett (2007) conducted a communication research about global strife, and its network interpretation result indicated that countries that are culturally similar tend to quarrel with members of divergent groups, which means cultural heterogeneity is an indicator that can predict clashes. According to this study, cultural differences may lead to conflicts or potential conflicts due to a lack of understanding between countries. Therefore, this study adopts cultural difference as a predictor of the different understandings of start-ups. Blei and Jordan (2003) developed the Latent Dirichlet Allocation (LDA) model, and it has resulted in many research models for huge datasets. Subsequently, these models are preferred for text analysis because of their simplicity in analysis and helpfulness in moderating the dimensions of data, and their ability to clarify semantic consistent topics (Mimno & McCallum, 2012). The Hierarchical Dirichlet Process model for news research and replies from readers was used to compare the documents of news topics outputted by the model to the public agenda, in terms of how people distribute and discuss news (Kim et al., 2014). These studies are theoretically and technically rooted in a study by Kok et al. (1999), who proposed an access where report crowding was tagged with media agenda issues, called 'traditional labelling'. One of outstanding features of the topic model is its joint usage with social science to explain trends. Bae et al. (2013) identified that topics in Twitter spread faster than in other media, using a topic model trained with the 2012 Korean presidential election dataset. Park and Lim (2014) conducted data analysis on Internet diplomacy using network analysis and topic modelling. Kim and Diesner (2015) performed an analysis of network trends that change over time.

Jiao et al. (2019) conducted research on enterprise innovation in new fields using data on high-tech enterprises in China and argued that, when collaborative innovation is more active, the growth of new fields can be accelerated. Jeong et al. (2019) visualised intellectual property right data refined through topic modelling. Their generative topographic mapping illustrated potential Research and Business Development areas with a concrete strategic framework that policy makers may consider.

Some scholars have used more sophisticated techniques for multilateral analysis and prediction. Mimno and McCallum (2012) suggest a Dirichlet-multinomial regression (DMR) topic analysis model, which is based on a log-linear prior on document-topic spread for observing features and argued that the DMR model could specify arbitrarily structured document features. DMR is an upstream topic model with a particularly useful method for incorporating arbitrary document features. When performing DMR topic modelling by applying parameters, it is expected that a wide range of analysis will be possible. Lee and Song (2020) suggest a generalised Dirichlet Multinomial Regression (g-DMR) topic model. The g-DMR model helps to calculate the distribution of topics by categorical metadata, where the trend of topic weight change according to the classification of categorical metadata is analysed in the form of continuous data. For example, by using g-DMR topic analysis, it is possible to analyse the trend of topic weight change in East Asian and North American media over time. The understanding of discourse can be enhanced through comparative analysis based on culture and geography. To examine these entrepreneurial RIS differences by country and region, the following research questions were presented:

(1) How much do media report on start-ups innovation in East Asia and North America differ, revealed by topic models?
(2) How different are the discourses on start-ups innovation in media reports from East Asia and North America in the 2000s and 2010s, analysed by Dirichlet-multinomial regression (DMR) topic analysis?
(3) How different are the discourses on start-ups innovation in media reports from East Asia and North America in the 2000s and 2010s, analysed by generalised Dirichlet-multinomial regression (g-DMR) theme analysis?

3. Materials and methods

3.1 Data crawling

Articles about start-ups were collected from LexisNexis (www.lexisnexis.com), which is a database providing an archive of various documents such as newspaper, magazines, and legal cases. The 'Major World Publication' section in LexisNexis, which provides articles in major newspapers and magazines, was selected, and the search word was 'start-up.' After deciding the search term, the search results were limited to English documents published from 2000 to 2019. The content of the example article below is little adapted, as shown below (Table 2).

3.2 Pre-processing

The data collected was pre-processed through the LexisNexis. First, all duplicate articles, articles without text content, and articles with only figures were removed. After data pre-processing, the total number of articles used for actual analysis was 51,938. The content of the title and body were combined into a single column for text analysis, and a continent column that divides the articles into East Asian regions and North American regions was inserted using information from the media, as shown in Table 3.

Table 2. News data example.

Title	Body	Source	Date
'SoftBank took loss in'	'Investments of IT powerhouse sold' (…)	The New York Times	Dec. 10, 2019

Table 3. News dataset.

Text	Source	Continent	Date
SoftBank took loss in, Investments of IT powerhouse sold (…)	The New York Times	North America	Dec. 10, 2019

The sentences in the text column in the table above were segmented into morpheme units for actual text mining. In this study, python's spacy library was used. Only nouns, verbs, adjectives, and adverbs were extracted from the entire data. Lemmatisation, which converts words into headwords, was performed while extracting the parts of speech (Figure 2).

Words that repeatedly appeared together were converted into bigrams and made into a single word. Based on the refined words, a list was constructed, and stop words were removed. Since most of the data used in this study were newspaper or magazine articles, the names of newspaper or magazine companies were inserted into the stop words list, and the English stop words list provided by the nltk package was also used. Finally, the thesaurus was unified into one. All the above pre-processing processes are summarised in Table 4.

After data pre-processing, the results showing the publisher and the number of articles, continent, and country were extracted and are shown in Table 5.

3.3 Modelling

The research was conducted using the DMR topic model and g-DMR Topic model. The DMR topic model was used to differentiate the start-up discourse between the two continents, which were East Asia and North America. G-DMR analysis was used to analyse the trend of topic weight change over time. The parameters of DMR analysis are described in Table 6, and the parameters of g-DMR analysis are described in Table 7.

4. Results

RQ1: How much do media report on start-ups innovation in East Asia and North America differ, revealed by topic models?

Eight topics were identified from the results. The number of topics was adjusted according to the analysis method. Due to the nature of topic model analysis, although the distribution of topics may have a similar context, they do not appear as completely identical topics. That is, DMR analysis and g-DMR analysis may show a similar topic distribution but are not composed of the same topic. In this research question, the topic distribution was examined with a focus on DMR topic analysis.

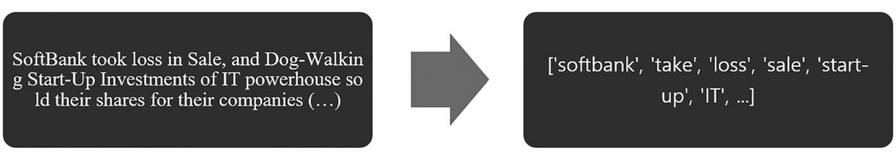


Figure 2. Text tokenising sample.

Table 4. Pre-processing process.

1. Tokenizing	2. Make bigram	3. Remove stop words	4. Unification of thesaurus
Extracts nouns, verbs, adjectives, and adverbs from sentences.	Words that appear together more than 1000 times are grouped into one word.	Remove words that have grammatical meaning or are not clear.	Unifying words such as abbreviations

Focusing on the DMR analysis results, the topic titles were found to be organised as follows. 'Physical space for start-ups', 'Beneficiary of innovation', 'Revenue sources', 'Emerging industry: mobility and energy', 'Innovation economy', 'Bank and fintech', 'Crowdfunding and VC', and 'Government policy'. After DMR analysis, the eight topics selected by us consist of detailed issues as follows. First, 'Physical space for start-ups' is related to shared office space, surrounding, and conditions for start-up business. 'Beneficiary of innovation' consists of smart device users, applications, and actors and beneficiaries from the start-up field. 'Revenue sources' is concerned with internet network, online advertisement, application, and games. 'Emerging industry: mobility and energy' is involved in electric vehicles, renewable energy, and robots. 'Bank and fintech' features subjects related to online payment systems. 'Crowdfunding and VC' is related to fund investment, venture capital, and new drug patents. 'Government policy' is composed of tax, law, rules, and political issues.

Table 5. Pre-processed news dataset.

Publisher	Count	Continent	Country
The New York Times	6,217	North America	U.S.A
China Daily	5,740	East Asia	China
The Boston Globe	3,828	North America	U.S.A
The Business Times Singapore	3,529	East Asia	Singapore
Daily Deal/The Deal	2,886	North America	U.S.A
South China Morning Post	2,445	East Asia	China
The Washington Post	2,221	North America	U.S.A
Ottawa Citizen	2,212	North America	Canada
Xinhua General News	2,036	East Asia	China
Los Angeles Times	1,977	North America	U.S.A
THE KOREA HERALD	1,796	East Asia	South Korea
The Toronto Star	1,650	North America	Canada
Korea Times	1,610	East Asia	South Korea
USA TODAY	1,256	North America	U.S.A
Wall Street Journal Abstracts	1,050	North America	U.S.A
Forbes	934	North America	U.S.A
Shenzhen Daily	906	East Asia	China
The Ottawa Citizen	760	North America	Canada
BusinessWorld	704	East Asia	Philippines
The Edge Singapore	591	East Asia	Singapore
Dow Jones Chinese Financial Wire	516	East Asia	China
Greater China Private Equity Review Daily	506	East Asia	China
SinoCast	460	East Asia	China
The Edge Malaysia	460	East Asia	Malaysia
The Japan Times	228	East Asia	Japan
People's Daily Online – English	194	East Asia	China
China Economic Review – Daily & Industry	179	East Asia	China
Beijing Review	177	East Asia	China
China Knowledge Newswire	124	East Asia	China
The Japan News	87	East Asia	Japan
Macau Business Daily	57	East Asia	China
ET Net	40	East Asia	China
Total	47,376		

Table 6. Parameters of DMR analysis.

Parameters	
Categorical variables	East Asia, North America
The number of topics	8
# of epochs	2000
Minimum document frequency of words.	400
Minimum collection frequency of words	800
Term weighting scheme	PMI[1]
Top words number to be removed	10
Random seed	2020

Table 7. Parameters of g-DMR analysis.

Parameters	
Categorical variables	East Asia, North America
The number of topics	8
# of epochs	1000
Minimum document frequency of words.	400
Minimum collection frequency of words	800
Top words number to be removed	10
Term weighting scheme	PMI
Random seed	2020
The degrees of Legendre polynomials for TDF(Topic Distribution Function)	3

RQ2: How different are the discourses on start-ups innovation in media reports from East Asia and North America in the 2000s and 2010s, analysed by Dirichlet-multinomial regression (DMR) topic analysis?

The trend of news reports by year between continents is as shown in Figure 3.

In the trend of news reports on start-up issues from 2000 to 2014, the number of reports in North America was higher than in East Asia. However, since 2015, the number of reports in East Asia has been increasing. In the East Asian region, it can be confirmed that the changes in the volume of news reports began in 2010.

Based on this, time variables (2000s, 2010s) in addition to the distinction between East Asia and North America were used as independent variables for the DMR analysis, which revealed the difference between the nominal variables more clearly, as shown in Table 8 below.

The table above shows the topic ratio. First, in the case of East Asia topics, in the 2000s, the themes of 'innovation economy' discourse and 'space' were central and in the 2010s, the agenda of the media shifted to topics related to 'emerging industry' and 'bank and fintech'. By contrast, in the case of North American topics, online network-oriented topics and government led innovation dominated in the 2000s and in the 2010s, themes related to life and crowdfunding and venture capital constituted the core elements. Labels of these topics can be examined in more detail from the words of each topic. The chart of the top 10 words for DMR analysis is shown below (Table 9).

Using the eight topics derived from the DMR topic analysis, a topic network analysis was conducted using gephi (https://gephi.org/) and has been used for network data mining and visualisation in many studies (Yang et al., 2017; Hussain et al., 2018). In this analysis, a study was conducted using the eigenvector centrality and weighted degree among the centrality indicators of semantic network analysis. Eigenvector centrality is calculated by applying the prestige index, which adds emphasis on the importance

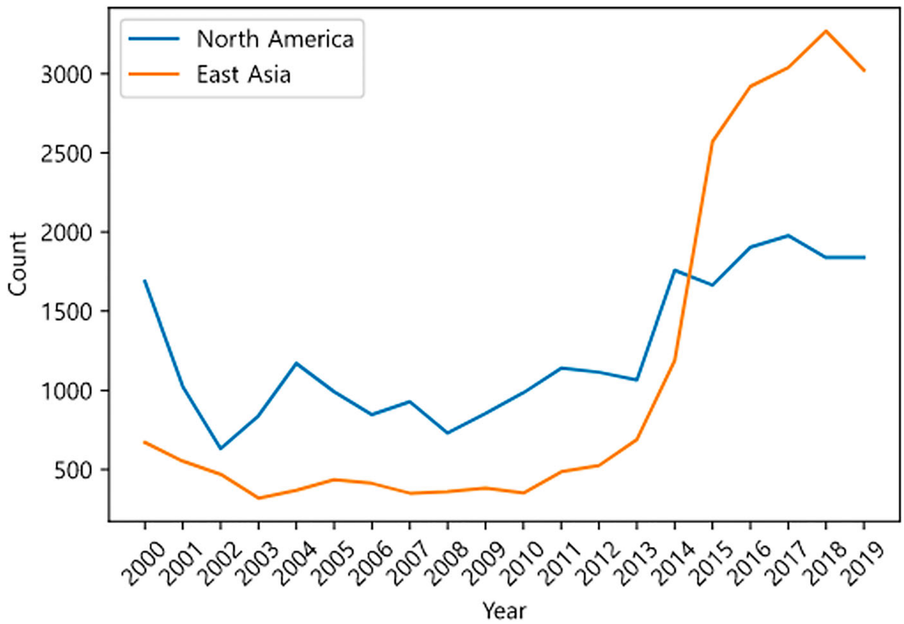


Figure 3. Number of news reports by year in North America and East Asia media.

Table 8. Result of DMR topic analysis.

	Topic1 Physical space for start-ups	Topic2 Beneficiary of innovation	Topic3 Revenue sources	Topic4 Emerging industry: mobility and energy	Topic5 Innovation economy	Topic6 Bank and fintech	Topic7 Crowd funding and VC	Topic8 Government policy
EA 2000s	22.63%	13.94%	5.34%	2.00%	**35.23%**	2.61%	14.60%	3.59%
EA 2010s	15.89%	6.08%	2.28%	**40.13%**	2.98%	16.66%	4.09%	11.86%
NA 2000s	6.10%	2.29%	**40.25%**	2.98%	16.71%	4.10%	11.90%	15.64%
NA 2010s	2.20%	**38.66%**	2.87%	16.05%	3.94%	11.43%	15.02%	9.81%

Note. Bold: The highest portion in each time period. Underlined: The second highest portion in each time period.

Table 9. Top 10 words of DMR analysis.

	Topic 1 Physical space for startups	Topic 2 Beneficiary of innovation	Topic 3 Revenue sources	Topic 4 Emerging industry: mobility and energy	Topic 5 Innovation economy	Topic 6 Bank and fintech	Topic 7 Crowdfunding and VC	Topic 8 Government policy
1	space	woman	internet	car	innovation	bank	fund	tax
2	job	student	user	vehicle	chinese	consumer	investor	state
3	office	school	software	energy	government	financial	venture	rule
4	property	entrepreneur	app	robot	country	loan	investment	law
5	city	man	ad	food	global	payment	capital	government
6	growth	child	network	electric	development	store	stock	employee
7	rate	family	computer	plant	industry	insurance	raise	federal
8	worker	team	game	driver	economy	online	deal	patent
9	rent	young	phone	bike	enterprise	fintech	round	political
10	economy	idea	video	oil	support	credit	drug	trump

of other nodes connected to a specific node to measure the impact at each node in the network. This method assumes that a particular node connected to a node with stronger centrality may be more influential than a node connected with pendants or isolates. The weighted degree of a node is based on the number of edges for a node and measures the weights of the nodes and edges that are connected to the node. This indicator scores higher as it is a topic located at the centre of the network connection structure. The weighted degree index was applied to the size of the node label, and analysed the edges by giving weights for building a semantic network, as shown in Table 10. The table below shows the values of the centrality according to each division of the eight topics derived from the DMR topic analysis, as shown in Table 11.

There may be differences between topics that are emphasised in the topic network and other topics that take up a lot of weight in DMR analysis. First of all, in the case of DMR analysis, if there were many documents related to the topic, the weight increases. However, even if the number of articles on a topic is small, the relative importance can be checked by the centrality score.

In the above semantic network, the larger the text size, the higher the weighted degree value and the sharper the colour of the edge, the higher the degree of connectivity. First, Topic 5 was at the centre of the network in both the 2000s and 2010s in East Asia and in the network structure; it can be seen that Topic 5 is strongly related to other topics. As a result of semantic network analysis, the discourse in North America in the 2000s is as follows. In the network structure, Topic 2 is at the centre of the network, and Topics 1, 8 and 5 are together form the core of the network. This network is also similar to the 2000s topic network.

Looking at the structure of the topic network in the North American region, differences appear in the network connection structure in the 2000s and 2010s. First of all, from the North American topic structure in the 2000s, Topic 8 is the most representative among network nodes, and topics such as Topics 2, 1, and 3 together contribute to a network structure. In this network structure, it can be seen that Topic 8 is still important when looking at the eigenvector value, indicating that the role of the government was important in the discourse on start-ups in the early 2000s. In addition, Topics 7 and 3 show strong connectivity in the network structure, indicating that investments such as venture capital have strong connectivity in areas where profits can be generated. The topic structure in the North American region has been showing complexities of connectivity among topics as it approaches the 2010s. In the 2010s, Topic 2 is at the centre of the network, while topics 8, 1, and 7 are emphasised together. In particular, when looking at the eigenvector value, topics 1, 2, 7, and 8 are emphasised. This change was interpreted as an emphasis on connectivity among individual topics in start-up innovation. Next, in DMR analysis, the contents of individual topics can be more clearly grasped through example articles on the top two topics. The example article extracted representative literature through the DMR analysis.

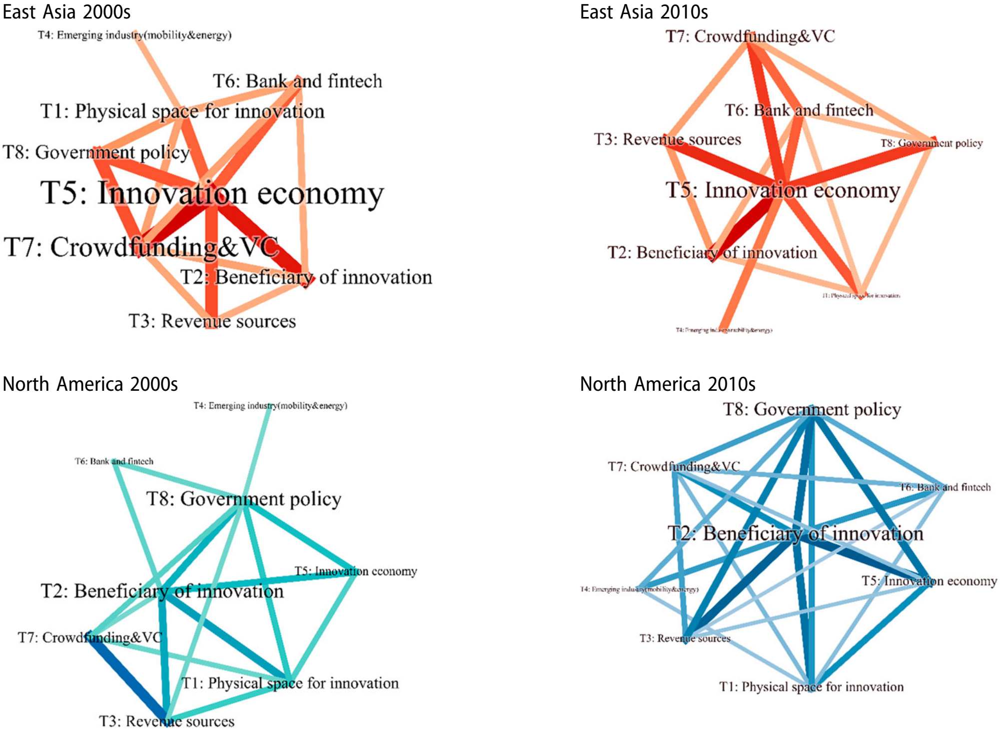
According to the results of DMR analysis in the 2000s in East Asia, the topic of 'innovation economy' had the greatest weight, and the issue of physical space related to start-ups increased. This provides the context on many start-ups becoming bankrupt during the economic crisis in the late 1990s in Korea, and jobs were recovered as they got close to the late 2000s. The data below show examples of articles on the top two topics derived from DMR analysis. According to the results of DMR analysis in the 2010s in

Table 10. Semantic network centrality scores.

Topic label	East Asia 2000s		East Asia 2010s		North America 2000s		North America 2010s	
	Eigen centrality	Weighted Degree	Eigen centrality	Weighted Degree	Eigen centrality	Weighted Degree	Eigen centrality	Weighted Degree
T1: Physical space for innovation	0.792298	3.317379	0.730873	3.032275	0.882881	3.612544	**1**	1.672946
T2: Beneficiary of innovation	0.744033	3.145778	0.696748	3.226361	0.958931	4.060381	**1**	**2.587648**
T3: Revenue sources	0.603646	2.841103	0.535507	2.957379	0.774799	3.492292	0.909004	1.455573
T4: Emerging industry (mobility & energy)	0.179788	2.204667	0.226115	2.323819	0.218614	2.256465	0.649893	0.759296
T5: Innovation economy	**1**	**4.486329**	**1**	**4.732348**	0.610267	2.987626	0.797497	1.558989
T6: Bank and fintech	0.7792	3.064688	0.857989	3.384785	0.422866	2.560931	0.797497	1.213575
T7: Crowdfunding&VC	**1**	3.866783	0.696748	3.250884	0.774799	3.34677	**1**	1.582858
T8: Government policy	0.616417	3.039409	0.730873	3.091943	**1**	**4.128102**	**1**	2.262526

Note. Bold: The highest portion in each time period. Underlined: The second highest portion in each time period.

Table 11. Inter-topic linkages.



East Asia, the ‘emerging industry: mobility and energy’ topic had the greatest weight. In addition, it can be seen that banking services that spread start-ups are being emphasised based on financial centres.

According to the results of DMR analysis in the 2000s in North America, the topic of ‘Revenue sources’ had the greatest weight. From the discourse of startups in the 2000s in North America, DMR analysis shows that small-scale business operators are playing the role of innovation actors in online platforms as well. The data below show example of articles on the top two topics derived from DMR analysis. According to the results of DMR analysis in the 2010s in North America, the topic of ‘beneficiary of innovation’ had the greatest weight. Meanwhile, from the discourses of North America in the 2010s, it can be seen that interest in beneficiaries who practically utilise advanced technologies is increasing, and that there is social interest in future foods.

RQ3: How different are the discourses on start-ups innovation in media reports from East Asia and North America in the 2000s and 2010s, analysed by generalised Dirichlet-multinomial regression (g-DMR) theme analysis?

The g-DMR topic model was analysed after setting the same number of topics as the DMR topic analysis. In this study, the trend of topic weight change over time was continuously analysed through g-DMR analysis, and the dynamics between the topics in the two regions were analysed. As a result of the analysis, the change in topic weight between East Asia and North America showed a change with the passage of time. Depending on the topic, there were cases where a specific region dominated the discourse or showed a similar direction, or the topic weight appeared alternately. In particular, due to the characteristics of the topic model, there was a noted difference in topic weight between DMR analysis and g-DMR analysis. In the case of the previous DMR analysis, since the change in topic weight was divided by 10 years, the meaning of individual topics was implied, and it is meaningful to show macroscopic trends, but it is insufficient to express detailed information. The Table 12 below shows the high word values as a result of g-DMR analysis.

Before examining the change trends of individual topics, this study examined how the proportion of the total topics had changed. To examine the trend of topic weight change in g-DMR analysis more clearly, the 20year period was divided into four sections, and the average value of topic weight in each section was calculated. Through this analysis, the dominant position of topics over time can be clearly identified.

As a result of the analysis, there was a clear difference in the topic weight of the East Asian region and the North American region, and there was a clear difference in the direction emphasised in the two cultures. First, the change in the proportion of topics in East Asia showed a direction in which the topic of government-led growth was greatly emphasised as time passed. In addition, the topics related to technology and technology users according to new innovations also gradually increased. Meanwhile, the proportion of topics related to finance, investment, and the energy industry showed a gradual decrease. This trend shows that the government's influence is greatly exercised in the initiative of innovation in the East Asian region. In the case of topic weighting in North America, topics on the beneficiaries of innovation and entrepreneurs are being emphasised, and topics on health care and energy fields have also been gradually strengthened. By contrast, crowdfunding issues show a declining trend over time. This trend of topic weight change in North America shows that innovation is being led by people, and it can be seen that digital technology-oriented innovation is spreading to various fields. The Table 13 below is the data on the topic weighting results of g-DMR analysis (Table 14).

First, in the case of Topic 1, topics related to banking and fintech appeared. In the DMR analysis, both the East Asian region and the North American region showed an increasing trend in the proportion of topics in the 2010s than in the 2000s. Topics in the East Asian region increased in the early 2000s, then gradually decreased after 2005, and then increased again after 2015. In the case of Topic 2, it shows the beneficiaries

Table 12. Top 10 words of g-DMR analysis.

	Topic 1 Bank and fintech	Topic 2 Beneficiary of innovation	Topic 3 Emerging industry: mobility	Topic 4 Revenue sources	Topic 5 Emerging industry: healthcare	Topic 6 Emerging industry: energy	Topic 7 Crowdfunding and VC	Topic 8 State-led development
1	Bank	Woman	Robot	Internet	Student	Food	Fund	Innovation
2	Financial	Job	Space	Consumer	Drug	Energy	Venture	Chinese
3	Stock	Employee	AI	User	School	Power	Capital	hong_kong
4	Tax	Team	Car	Online	Patient	Car	Round	Government
5	Share	Entrepreneur	Uber	App	Medical	Plant	Investor	Development
6	Loan	Family	City	Mobile	Research	Vehicle	Investment	Country
7	Fintech	Man	Bike	Game	Education	Oil	Deal	Economy
8	Price	Age	Driver	Video	University	Electric	Raise	Global
9	Government	Community	Vehicle	Store	Cancer	Production	Software	Industry
10	State	Worker	Office	Phone	Health	Water	Partner	Economic

Table 13. Topic proportion ratio of g-DMR.

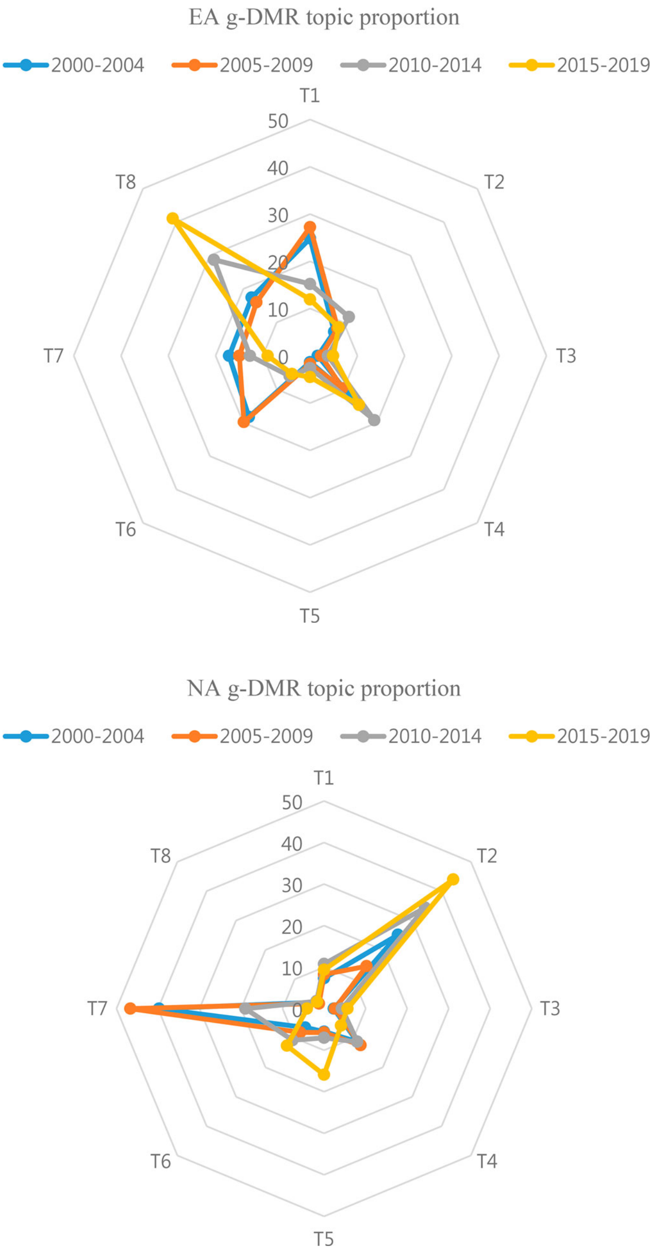
	Year	Topic 1 Bank and fintech	Topic 2 Beneficiary of innovation	Topic 3 Emerging industry: mobility	Topic 4 Revenue sources	Topic 5 Emerging industry: healthcare	Topic 6 Emerging industry: energy	Topic 7 Crowdfunding and VC	Topic 8 State-led development
EA	2000–2004	24.86	7.22	1.618	12.06	1.322	18.294	17.124	17.46
	2005–2009	27.234	8.208	2.336	9.682	1.7	19.8	14.958	16.034
	2010–2014	15.204	11.626	3.782	19.234	2.59	6.076	12.706	28.74
	2015–2019	11.932	8.526	4.878	14.67	4.488	5.412	8.96	41.09
NA	2000–2004	7.322	25.134	2.326	11.254	5.548	6.428	39.744	2.208
	2005–2009	8.25	14.46	2.576	12.456	5.748	8.106	46.632	1.732
	2010–2014	10.754	34.362	4.286	11.304	7.026	10.752	18.93	2.552
	2015–2019	9.35	43.976	5.64	5.774	15.922	12.628	4.11	2.354

of innovative technology and discourse on education and the dominance of topics in this discourse is emphasised mainly in North America. As smart technology becomes more common, issues related to education and life of related technologies appear to be highlighted. In the case of Topic 3, the topic of innovation in the mobility industry was emphasised and considering the flow of topic weight, it can be said that the discourse in North America and the discourse in East Asia are emphasised in a similar way. This trend of increasing topics has been gradually strengthened with the development of artificial intelligence technology since 2007 and confirms that innovation in the mobility field is being emphasised in both East Asia and North America. In the case of Topic 4, the topic of the source of actual revenue is emphasised, in particular, the Internet and the smartphone application ecosystem. As for the proportion of topics in East Asia, Internet issues gradually stabilised in the early 2000s, but since 2007, as smartphones spread, news reports became more emphasised. By contrast, topics in North America are relatively less volatile than in East Asia, and the proportion of topics tends to decrease. In the case of Topic 5, it is a topic in which the health care industry group has strong issues. Topics related to healthcare are being emphasised especially in North America, and the upward trend has been increasing since 2015 and appears to be due to the development of artificial intelligence technology since 2015. In the case of Topic 6, topics related to food and energy are being emphasised. First, the proportion of topics related to the subject in North America is steadily increasing. By contrast, in the case of East Asia, it increased rapidly in the early 2000s, but after 2005, the proportion of topics related to the topic decreased. In the case of Topic 7, content related to capital investment and venture capital is emphasised. Looking at the topic flow in North America, topics related to investment and funding increased in the early 2000s and then gradually declined, and in the case of East Asia, they did not show a significant impact. In the case of Topic 8, the topic of government-led industries is being emphasised and showed that the East Asian region is more emphasised than the North American region. The Table 15 below is the data on the weight of individual topics as a result of g-DMR analysis.

5. Discussion

As a result of the analysis, different innovation factors were emphasised in the discourse of startups in East Asia and startups in North America, and it was found that there is a

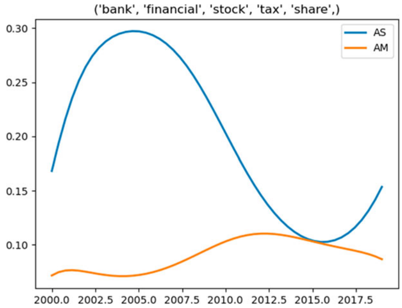
Table 14. Topic proportion graph of g-DMR.



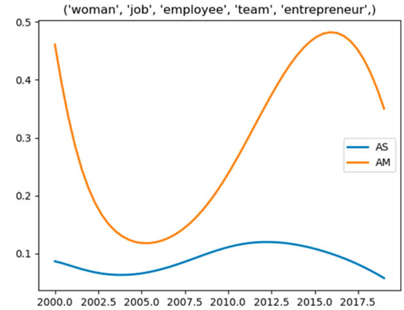
difference according to maturity in the dimension of entrepreneurial RIS. In the case of North America, since the establishment of social infrastructure is sufficient, the actors who constitute a start-up were emphasised, whereas in the East Asia region, the discourse about the government and large corporations for the establishment of social infrastructure was emphasised. According to Yoon et al. (2015), in the case of North America, it belongs to the category of Mature entrepreneurial RIS, and it was seen that the formation

Table 15. g-DMR analysis result.

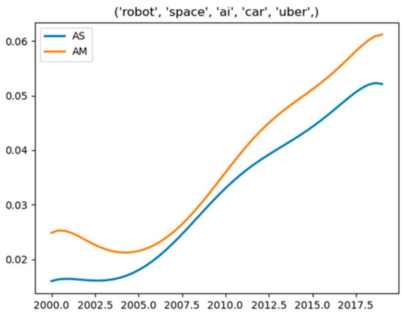
Topic1: Bank and fintech



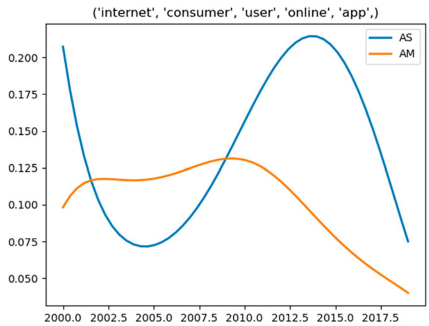
Topic2: Beneficiary of innovation



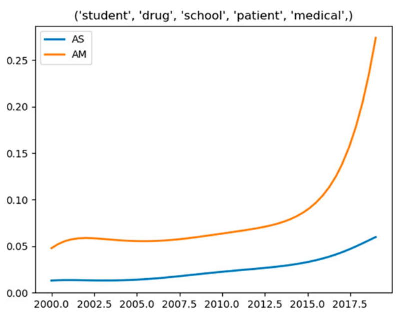
Topic3: Emerging industry: mobility



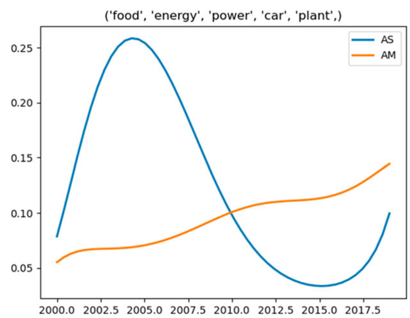
Topic4: Revenue sources



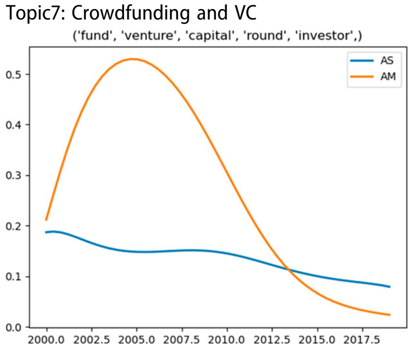
Topic5: Emerging industry: healthcare



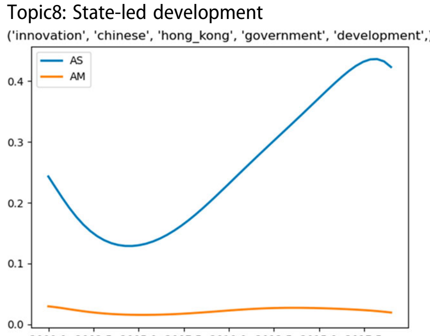
Topic6: Emerging industry: energy



Topic7: Crowdfunding and VC



Topic8: State-led development



of startups occurs naturally. In particular, he argued that startups are formed by individual entities such as entrepreneurs and VCs. On the other hand, in the case of the East Asian region, it belongs to the category of Still-evolving entrepreneurial RIS, and the formation of start-ups appears to be led by the government.

First of all, in the discourse of startups in North America in the 2000s, as a result of DMR analysis, technology-oriented factors such as internet, software, and user were emphasised. This is interpreted as innovation in the aspect of R&D. In addition, it was confirmed that the element of innovation itself was emphasised. And as a result of g-DMR analysis, discourse on crowdfunding and VC was emphasised. As a result of DMR analysis, the discourse on startups in North America in 2010 was changed and emphasised by individual entities constituting startups such as entrepreneur, team, young, and idea. And as a result of g-DMR analysis, the topic of Beneficiary of Innovation was emphasised. From the perspective of entrepreneurial RIS, when looking at the discourse of startups in North America, the government-led innovation system was built, and it was confirmed that the individual actors of innovation according to the mature entrepreneurial RIS were highlighted as a result of news analysis.

Meanwhile, in the discourse of startups in East Asia in the 2000s, the DMR analysis showed that government-led innovation such as chinese, government, country, and support was emphasised. In addition, physical space-oriented startups such as space, property, and rent were emphasised, which is different from the emphasis on software-centered discourse in North America. And as a result of g-DMR analysis, topics on basic facilities such as financial and plant were emphasised. In the 2010s, as a result of DMR analysis, the discourse on startups in East Asia changed the discourse of innovation centered on large industries such as car, robot, energy, and plant, and financial support such as bank, loan, and credit helped innovative growth. emphasised. And as a result of g-DMR analysis, it was confirmed that the discourse on state-led development was dominant. In this aspect, as a key factor leading startups in East Asia at the level of still-evolving entrepreneurial RIS, the factors for the composition of social infrastructure are emphasised, and conglomerates and the government are continuously emphasised as agents of action.

6. Conclusion

6.1 Key findings

As a result of the analysis, the discourse on startups in East Asia and those in North America revealed differences according to Entrepreneurial RIS. As a result of the topic network analysis, it was found that the interrelationship between the topics constituting the startup innovation factors in East Asia and North America was high. In order to examine the innovation factors of startups, we focused on the results of DMR and g-DMR analysis.

From the 2000s to the 2010s, the discourse on startups in the East Asian region, which belongs to the still-evolving entrepreneurial RIS geopolitically, is emphasised by the government and large businesses centered on large businesses, and financial support for this is emphasised centered on the existing financial sector. These analysis results suggest that entrepreneurial RIS of startups in East Asia has not yet fully entered the mature stage.

On the other hand, in the North American startup discourse, which belongs to the geographically mature entrepreneurial RIS, innovation has emerged centered on software in the 2000s, and as the 2010s move on, the discourse changes centered on actors and Venture Capitals that make up startups. These analysis results clearly inform that the entrepreneurial RIS of startups in North America is in the mature stage.

Based on this analysis result, in order for entrepreneurial RIS in East Asia to move to a mature stage, the composition of the startup innovation system must be changed in terms of companies and governments. First of all, in terms of companies, innovation centered on large enterprises is limited in terms of the derivation of innovative ideas. Accordingly, it is necessary to change into a system that can be created naturally by companies such as SMEs.

In addition, from the government's perspective, it is necessary to change the direction of the system in a form that can form various start-ups by expanding the range of support for start-up support, away from the public relations method of nurturing start-ups that fit the theme set by the government. Considering that government is one of the key components of a start-up ecosystem (Tripathi et al., 2019; Ziakis et al., 2022), governments should understand the mechanisms for creating and disseminating innovation, to promote and not impede business activities (Feld, 2020). For example, tax incentives or acceleration of starting processes may be good strategies for enhancing the effectiveness of operation of start-up ecosystem (Ziakis et al., 2022).

This study is meaningful in that it is an empirical study that reveals that there is a substantial difference between the East Asian region and the North American region in terms of entrepreneurial RIS. Also, at the practical level, the study is meaningful in terms of suggesting the direction of startup incubation systems in countries that are seen to exist in East Asia and other still-evolving entrepreneurial RIS.

6.2 Limitations and suggestions for future research

Despite this research has some contributions to start-up innovation area, there are some issues that should be addressed in the future. First, the data employed in this study are mainly from USA and China. Thus, future studies can collect and analyse more data sourced from other countries in North America and East Asia. Second, in terms of technical approaches, future research can employ sentiment analysis that has been widely applied in past research (see Kim et al., 2022) to analyse start-up related news articles.

Note

1. Pointwise mutual information (PMI) is a measure of the interdependence between two variables in information theory. In this study, PMI was used to calculate the interdependence between two terms.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Kyeo Re Lee is a SW collaboration professor at Department of Software, Hanyang University ERICA, Ansan, South Korea. His research focuses on computational social sciences, human-AI interaction, metaverse and data science.

Jang Hyun Kim is a professor at Department of Interaction Science, and a joint affiliate faculty at Department of Applied Artificial Intelligence, Sungkyunkwan University and his research focuses on social/semantic data analysis, social media, and future media.

Jaeyeon Jang is a master at Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, South Korea. Her reaserch intersts include data science, social network anlaysis and human-AI interaction.

Jeewoo Yoon is a Ph.D. student at Department of Applied Artificial Intelligence in Sungkyunkwan University, Seoul, South Korea. His research interests include multimodal machine learning and affective computing.

Dongyan Nan is a Ph.D. candidate at Department of Human-Artificial Intelligence Interaction/ Department of Interaction Science, Sungkyunkwan University. His research focuses on Human-AI Communication, Consumer Experience, Business Informatics, Computational Social Science, Metaverse, and Network Analysis. He has authored 20 papers in major conferences and journals such as Technological Forecasting and Social Change, Information Processing and Managment, and Journal of Organizational and End User Computing.

Yonghwan Kim is working for Naver and PhD in Psychology. His research focuses on data science, metaverse, and social media.

Byungjun Kim is a research associate in Center for Digital Humanities and Computational Social Sciences, Korea Advanced Institute of Science and Technology, Daejeon, South Korea. His research interests include digital humanities, computational social sciences, and natural language processing.

ORCID

Kyeo Re Lee http://orcid.org/0000-0002-1621-8659
Jang Hyun Kim http://orcid.org/0000-0001-7750-2664
Jaeyeon Jang http://orcid.org/0000-0001-5852-7751
Jeewoo Yoon http://orcid.org/0000-0002-9067-8653
Dongyan Nan http://orcid.org/0000-0001-7811-0857
Yonghwan Kim http://orcid.org/0000-0003-4711-6011
Byungjun Kim http://orcid.org/0000-0001-9925-9343

References

Ahn, S., Lee, K. H., & Kim, K. (2018). Assessing the relationship between technological innovation regulation and performance of small and medium enterprises in Korea: Focusing on the effects of individual regulations in a firm's business life cycle. *Journal of Social Science*, *57*(2), 159–195. https://doi.org/10.22418/JSS.2018.12.57.2.159

Avermaete, T., Viaene, J., Morgan, E. J., Pitts, E., Crawford, N., & Mahon, D. (2004). Determinants of product and process innovation in small food manufacturing firms. *Trends in Food Science & Technology*, *15*(10), 474–483. https://doi.org/10.1016/j.tifs.2004.04.005

Bae, J. H., Son, J. E., & Song, M. (2013). Analysis of twitter for 2012 South Korea presidential election by text mining techniques. *Journal of Intelligence and Information Systems*, *19*(3), 141–156. https://doi.org/10.13088/jiis.2013.19.3.141

Blei, D. M., & Jordan, M. I. (2003, July). Modeling annotated data. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 127–134).

Bogers, M., Chesbrough, H., & Moedas, C. (2018). Open innovation: Research, practices, and policies. *California Management Review*, *60*(2), 5–16. https://doi.org/10.1177/0008125617745086

Boston Consulting Group. (2020). *The most innovative companies 2020 the serial innovation imperative*. BCG.

Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press.
Chong, D., & Druckman, J. N. (2007). A theory of framing and opinion formation in competitive elite environments. *Journal of Communication*, *57*(1), 99–118. https://doi.org/10.1111/j.1460-2466.2006.00331_3.x.
Cohen, B. (1963). *The press and foreign polcy*. Harcourt.
Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128–152. https://doi.org/10.2307/2393553
Colombelli, A., Grilli, L., Minola, T., & Mrkajic, B. (2020). To what extent do young innovative companies take advantage of policy support to enact innovation appropriation mechanisms? *Research Policy*, *49*(10), 103797. https://doi.org/10.1016/j.respol.2019.05.006
Cooke, P., & Leydesdorff, L. (2006). Regional development in the knowledge-based economy: The construction of advantage. *The Journal of Technology Transfer*, *31*(1), 5–15. https://doi.org/10.1007/s10961-005-5009-3
Cooke, P., Uranga, M. G., & Etxebarria, G. (1997). Regional innovation systems: Institutional and organisational dimensions. *Research Policy*, *26*(4–5), 475–491. https://doi.org/10.1016/S0048-7333(97)00025-5
Feezell, J. T. (2018). Agenda setting through social media: The importance of incidental news exposure and social filtering in the digital era. *Political Research Quarterly*, *71*(2), 482–494. https://doi.org/10.1177/1065912917744895
Feld, B. (2020). *Startup communities: Building an entrepreneurial ecosystem in your city*. John Wiley & Sons.
Freeman, J., & Engel, J. S. (2007). Models of innovation: Startups and mature corporations. *California Management Review*, *50*(1), 94–120. https://doi.org/10.2307/41166418
Guckenbiehl, P., de Zubielqui, G. C., & Lindsay, N. (2021). Knowledge and innovation in start-up ventures: A systematic literature review and research agenda. *Technological Forecasting and Social Change*, *172*, 121026. https://doi.org/10.1016/j.techfore.2021.121026
Hinings, B., Gegenhuber, T., & Greenwood, R. (2018). Digital innovation and transformation: An institutional perspective. *Information and Organization*, *28*(1), 52–61. https://doi.org/10.1016/j.infoandorg.2018.02.004
Huang, H. C., Lai, M. C., & Lo, K. W. (2012). Do founders' own resources matter? The influence of business networks on start-up innovation and performance. *Technovation*, *32*(5), 316–327. https://doi.org/10.1016/j.technovation.2011.12.004
Hussain, S., Muhammad, L. J., & Yakubu, A. (2018). Mining social media and DBpedia data using gephi and R. *Journal of Applied Computer Science & Mathematics*, *12*(1), 14–20. https://doi.org/10.4316/JACSM.201801002
Iammarino, S. (2005). An evolutionary integrated view of regional systems of innovation: Concepts, measures and historical perspectives. *European Planning Studies*, *13*(4), 497–519. https://doi.org/10.1080/09654310500107084
Jeong, Y., Park, I., & Yoon, B. (2019). Identifying emerging research and business development (R&BD) areas based on topic modeling and visualization with intellectual property right data. *Technological Forecasting and Social Change*, *146*, 655–672. https://doi.org/10.1016/j.techfore.2018.05.010
Jiao, H., Yang, J., Zhou, J., & Li, J. (2019). Commercial partnerships and collaborative innovation in China: The moderating effect of technological uncertainty and dynamic capabilities. *Journal of Knowledge Management*, *23*(7), 1429–1454. https://doi.org/10.1108/JKM-10-2017-0499.
Jugend, D., Fiorini, P. D. C., Armellini, F., & Ferrari, A. G. (2020). Public support for innovation: A systematic review of the literature and implications for open innovation. *Technological Forecasting and Social Change*, *156*(February), 119985. https://doi.org/10.1016/j.techfore.2020.119985
Kang, S. (2019). A study on the structural relationships among R&D resources, R&D collaboration, policy support, and operational performance — Focusing on Korean health industry clusters. *Management & Information Systems Review*, *38*(2), 225–240. https://doi.org/10.29214/damis.2019.38.2.012

Kim, J., & Diesner, J. (2015). The effect of data pre-processing on understanding the evolution of collaboration networks. *Journal of Informetrics*, *9*(1), 226–236. https://doi.org/10.1016/j.joi.2015.01.002

Kim, J. H., & Barnett, G. A. (2007). A structural analysis of international conflict: From a communication perspective. *International Interactions*, *33*(2), 135–165. https://doi.org/10.1080/03050620701277764

Kim, J. H., Jung, H. S., Park, M. H., Lee, S. H., Lee, H., Kim, Y., & Nan, D. (2022). Exploring cultural differences of public perception of artificial intelligence via big data approach. In J. H. Kim, H. S. Jung, M. H. Park, S. H. Lee, H. Lee, Y. Kim, & D. Nan (Eds.), *International conference on human-computer interaction* (pp. 427–432). Springer.

Kim, J. H., Su, T. Y., & Hong, J. (2007). The influence of geopolitics and foreign policy on the US and Canadian media: An analysis of newspaper coverage of Sudan's Darfur conflict. *Harvard International Journal of Press/Politics*, *12*(3), 87–95. https://doi.org/10.1177/1081180X07302972

Kim, Y. (2005). Technological collaboration linkages and the innovation output in small and medium-sized firms: A study on the moderating effects of absorptive capacity. *Korean Management Review*, *34*(5), 1365–1390.

Kim, Y., Kim, S., Jaimes, A., & Oh, A. (2014). A computational analysis of agenda setting. Proceedings of the 23rd International Conference on World Wide Web – WWW '14 Companion. doi:10.1145/2567948.2577379.

Kok, Y. H., Goh, A., & Holaday, D. (1999). Agenda: A tool for agenda setting research. *Information Services & use*, *19*(4), 265–276. https://doi.org/10.3233/ISU-1999-19403

Lau, A. K., & Lo, W. (2015). Regional innovation system, absorptive capacity and innovation performance: An empirical study. *Technological Forecasting and Social Change*, *92*, 99–114. https://doi.org/10.1016/j.techfore.2014.11.005

Lee, M., & Song, M. (2020). *Incorporating citation impact into analysis of research trends*. Scientometrics.

Lew, Y. K., & Park, J. Y. (2021). The evolution of N-helix of the regional innovation system: Implications for sustainability. *Sustainable Development*, *29*(2), 453–464. https://doi.org/10.1002/sd.2143

Majamaa, W. (2008). *The 4th P-People-in urban development based on Public-Private-People Partnership*. Teknillinen korkeakoulu.

Miller, M. M. (1997). Frame mapping and analysis of news coverage of contentious issues. *Social Science Computer Review*, *15*(4), 367–378. https://doi.org/10.1177/089443939701500403

Mimno, D., & McCallum, A. (2012). Topic models conditioned on arbitrary features with dirichlet-multinomial regression. arXiv:1206.3278.

Nah, S., McNealy, J., Kim, J. H., & Joo, J. (2020). Communicating artificial intelligence (AI): Theory, research, and practice. *Communication Studies*, *71*(3), 369–372. https://doi.org/10.1080/10510974.2020.1788909.

Nguyen, V. (2015). *Guided Probabilistic Topic Models for Agenda-setting and Framing*. (Ph.D. dissertation) University of Maryland, College Park, USA.

Park, S. J., & Lim, Y. S. (2014). Information networks and social media use in public diplomacy: A comparative analysis of South Korea and Japan. *Asian Journal of Communication*, *24*(1), 79–98. https://doi.org/10.1080/01292986.2013.851724

Son, S. H., & Hong, C. Y. (2019). Potential conflict resolution methods by conflict types in interisland marine transportation planning: Udo-island ferry operation case study. *Korean Association Of Cadastre Information*, *21*(2), 120–139. https://doi.org/10.46416/JKCIA.2019.08.21.2.120

Spender, J. C., Corvello, V., Grimaldi, M., & Rippa, P. (2017). Startups and open innovation: A review of the literature. *European Journal of Innovation Management*, *20*(1), 4–30. https://doi.org/10.1108/EJIM-12-2015-0131

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, *15*(6), 285–305. https://doi.org/10.1016/0048-7333(86)90027-2

Tripathi, N., Oivo, M., Liukkunen, K., & Markkula, J. (2019). Startup ecosystem effect on minimum viable product development in software startups. *Information and Software Technology*, *114*, 77–91. https://doi.org/10.1016/j.infsof.2019.06.008

Yang, J., Cheng, C., Shen, S., & Yang, S. (2017). Comparison of complex network analysis software: Citespace, SCI 2 and gephi. In J. Yang, C. Cheng, & S. Yang (Eds.), *2017 IEEE 2nd international conference on big data analysis (ICBDA)* (pp. 169–172). IEEE.

Yoon, H., Yun, S., Lee, J., & Phillips, F. (2015). Entrepreneurship in East Asian regional innovation systems: Role of social capital. *Technological Forecasting and Social Change*, *100*, 83–95. https://doi.org/10.1016/j.techfore.2015.06.028

Ziakis, C., Vlachopoulou, M., & Petridis, K. (2022). Start-up ecosystem (StUpEco): A conceptual framework and empirical research. *Journal of Open Innovation: Technology, Market, and Complexity*, *8*(1), 35. https://doi.org/10.3390/joitmc8010035