

Topic Mining and Evolution Analysis of Green Logistics Using Latent Dirichlet Allocation Based Topic Modeling Technology

마서준디(MA RUIJUNDI)

인하대학교 물류전문대학원 대학원생

김용진 교수

인하대학교 아태물류학부

Abstract

Background: Green logistics is the main development trend of modern logistics. It is necessary to analyze the evolution of green logistics from a macro perspective.

Objective: This study aims to analyze the evolution of green logistics by using the Latent Dirichlet Allocation (LDA) topic model based on the article data.

Methodology: Latent Dirichlet Allocation, topic modeling, text similarity

Data: 3437 articles in web of science from 1993 to 2021

Contribution: This study analyzes the evolution of green logistics based on article data and extends the LDA topic model to a certain extent.

CONTENTS

-
- 01. Introduction**
 - 02. Literature review**
 - 03. Research method**
 - 04. Experiment and results**
 - 05. Conclusions**
 - 06. Reference**

Green logistics is the main development trend of modern logistics. As data of the International Energy Agency (IEA) shows that the **CO₂ emissions** of the **transportation sector** is 8258 million tones in 2018, which is **24.64%** of entire world. Therefore, many scholars focus on how to balance the contradiction between maximise profits and the reduction of the harmful impact on the environment.

Although the discussion on green logistics has decreased recently, there is no doubt that the development of green logistics will still be the **mainstream** of logistics. It is necessary to analyze the evolution of green logistics based on article data from a macro perspective.

1 Introduction

This study systematically divides the evolution of green logistics into three stages based on technology life cycle theory.

What's more, this study calculates the number of topics in each development stage and makes sure the meaning of each topic by using using the Latent Dirichlet Allocation (LDA) topic model based on the article data.

Finally, this paper visualizes the development roadmap of green logistics and analyzes the evolution among three stages.

2 Literature review-Green logistics

Based on the literature analysis in the field of green logistics, we divide the current research on green logistics into six parts, including **green purchasing**, **green warehousing**, **green production**, **green transportation**, **green delivery** and **reverse logistics**.

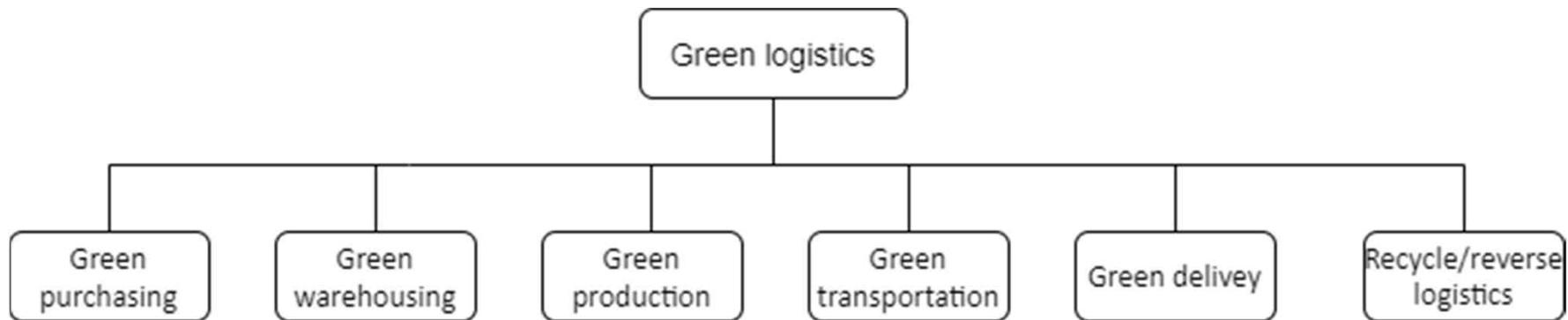


Fig 1. The main parts of green logistics

2 Literature review-Green logistics

Table 1. Literature review of green logistics

	Dimension	Literature
Green/sustainable purchasing	Green purchasing policy (GPP)	(Miyamoto, et al., 2020) (Foo M Y, et al., 2019) (Khoiruman M, Haryanto A T., 2017) (Murray J G., 2000)
	Green purchase intentions	(Marvi, et al., 2020) (Rizwan M, et al., 2014) (Chen K, Deng T, 2016) (Khaola P P, et al., 2014)
	Factors influencing green purchase intention	(Zhuang, et al., 2021) (Patel J D, et al., 2020)
Green/sustainable warehousing	Green warehouse management	(Bank R, Murphy R, 2021) (Chen X, et al., 2016)
	Environmental impact of warehouse building	(Huang J, Gurney K R, 2016) (Ries J M, 2017)
	Energy saving in warehousing	(Freis J, et al., 2016) (Burinskiene A, et al., 2018)
Green/sustainable production	Smart manufacturing	(Straka M, et al., 2021) (Mittal S, et al., 2019)
	Real-time production logistics	(Chen Z, Bidanda B, 2019) (Qu T, et al., 2016)
	Green packaging	(Wandosell, et al., 2021) (Moustafa H, et al., 2019)

2 Literature review-Green logistics

	Dimension	Literature
Green/sustainable transportation	Assessment of transportation system	(Naganathan, et al., 2017) (Kelarestaghi K B, 2019)
	Green transportation strategy and challenge	(Fang X, et al., 2020) Marcucci, et al., 2013)(Guo A, 2017)
	Key determinants of sustainable transportation	(Sayyadi R, Awasthi A, 2018) (Velazquez L, 2015)(Ahmad S, de Oliveira J A P., 2016)
Green/sustainable delivery	Optimal delivery strategy	(Li H C, 2015) (Silva A S,et al., 2014)
	Last mile delivery	(Jiang, et al., 2020) (Giret A, et al., 2018)
	Green vehicle routing problem	(Majidi, Setareh, et al.,2017) (Liu G, et al., 2020)
	Assignment of green delivery	(Pani A, et al., 2020) (Kong L, et al., 2018)
Green/sustainable waste recycle/reverse logistics	Green recycle strategy	(Sellitto M A, de Almeida F A, 2019)
	Structure of reverse logistics	(Khor K S, Udin Z M., 2013) (Agrawal S, et al., 2015)
	Critical factors	(Yang,et al., 2020) (De Campos,et al., 2020)
	Optimal model	(Alfonso-Lizarazo,et al., 2013) (Roghanian E, Pazhoheshfar P., 2014)

2 Literature review-LDA topic model

Table 2. Literature review of LDA topic model

Authors	Year	Research object	Main methodology	Data
Zhang H, Daim T, et al.	2021	Blockchain	LDA topic model; Cosine similarity	Patent
Du Y J, et al.	2020	Micro-blog features	LDA; Hot topic tracking (HTT) algorithm	Information from Micro-blog posts
Bastani K, et al.	2019	Consumer complaint narratives	LDA topic model;	CFPB consumer complaint database
Roque C, et al.	2019	Roadside safety	Text mining; LDA topic model	Road Safety Inspections(RSI)reports
Ding S, et al.	2019	Diabetic complication	similarity enhanced LDA model	Medical records dataset

2 Literature review-Research gap

Based on the literature review, many scholars have done a lot of research in the field of green logistics and LDA topic model, but this research **proposes to apply LDA topic model in the field of green logistics**, and put forward opinions on the evolution of green logistics. This study **supplements** the evolution of green logistics.

3 Research method

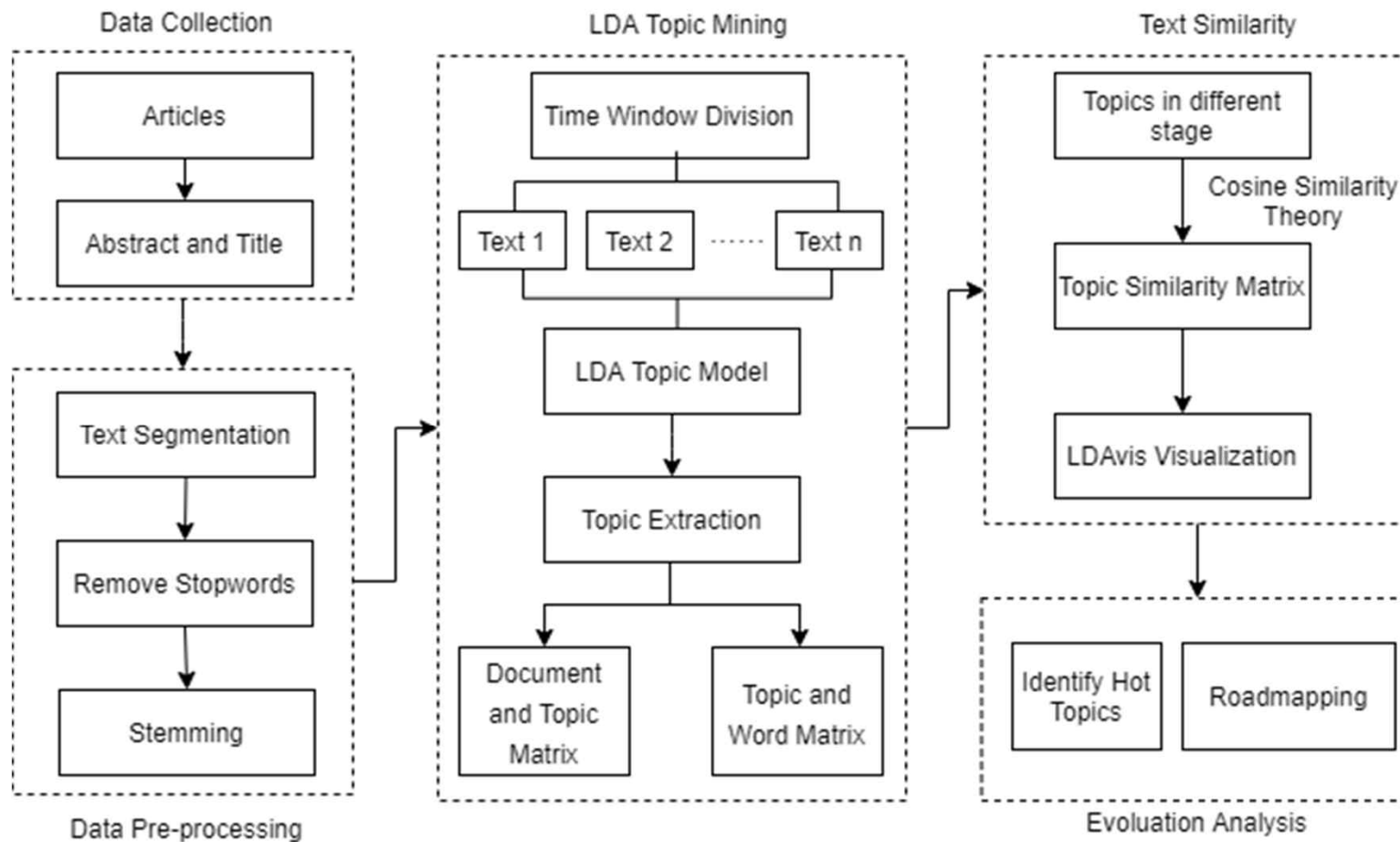


Fig 2. LDA-based research method flow

3 Research method-Topic modeling

- In machine learning and natural language processing, topic models are generative models, which provide a **probabilistic framework**.
- Topic modeling is a statistical model for discovering abstract topics in a series of documents.

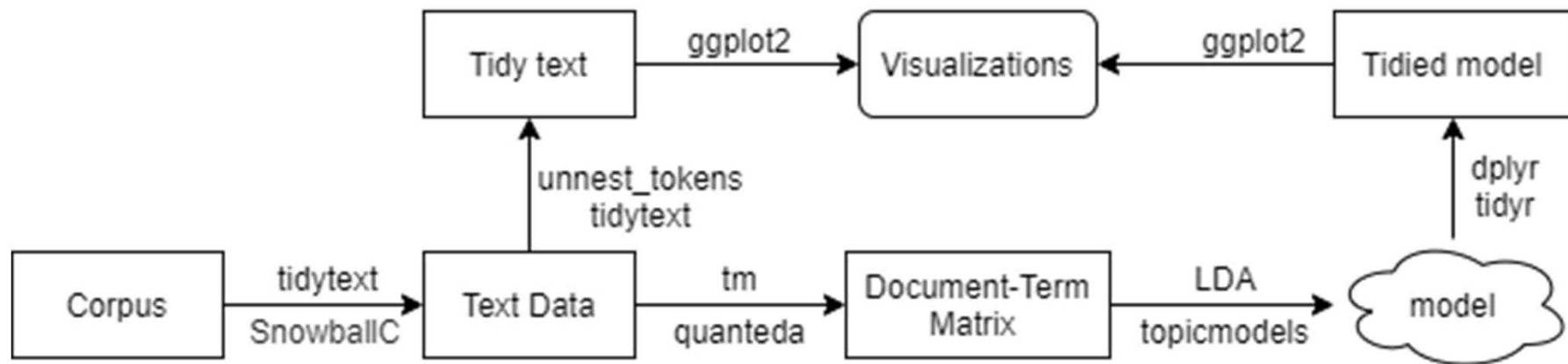


Fig 2. A flowchart of topic modeling based on R programming

3 Research method-LDA

Latent Dirichlet allocation (LDA) is a generative model for topic modeling. The LDA topic model extracts topics from text, which is a document topic generation model, which **contains a three-layer structure of words, topics, and documents.**

There are two principles of LDA:

- Every document is a mixture of topics
- Every topic is a mixture of words

3 Research method-LDA

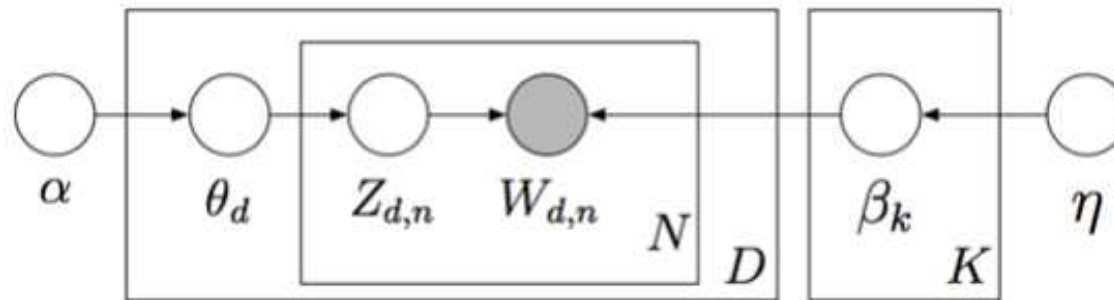


Fig 3. Structure for LDA topic model

Notation:

D: All documents in the corpus

N: The number of words in each document

K: Total number of topics

α : Proportion parameter

η : Topic parameter

θ_d : Distribution of topics on document d

β_k : Distribution of words on topic k

$Z_{d,n}$: The subject of the n th word in document d

$W_{d,n}$: The n th word in document d

Source: Introduction to Probabilistic Topic Models by David M. Blei

3 Research method-Text similarity

Text similarity calculation refers to comparing the similarity of two or more entities using a method that yields a specific quantified similarity value. This study uses **cosine similarity** to give similarity between two topics in terms of their subject matter.

Given two attribute vectors, A and B, the remaining string similarity θ is given by the dot product and the vector length as follows:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

The similarity of the results ranges from -1 for complete opposites, to 1 for complete sameness, 0 for orthogonality or decorrelation, and intermediate values for intermediate similarity or dissimilarity.

4 Experiment and results

Data source

This study uses article data to perform the whole development process of green logistics, and uses the **Web of Science** database as its data source.

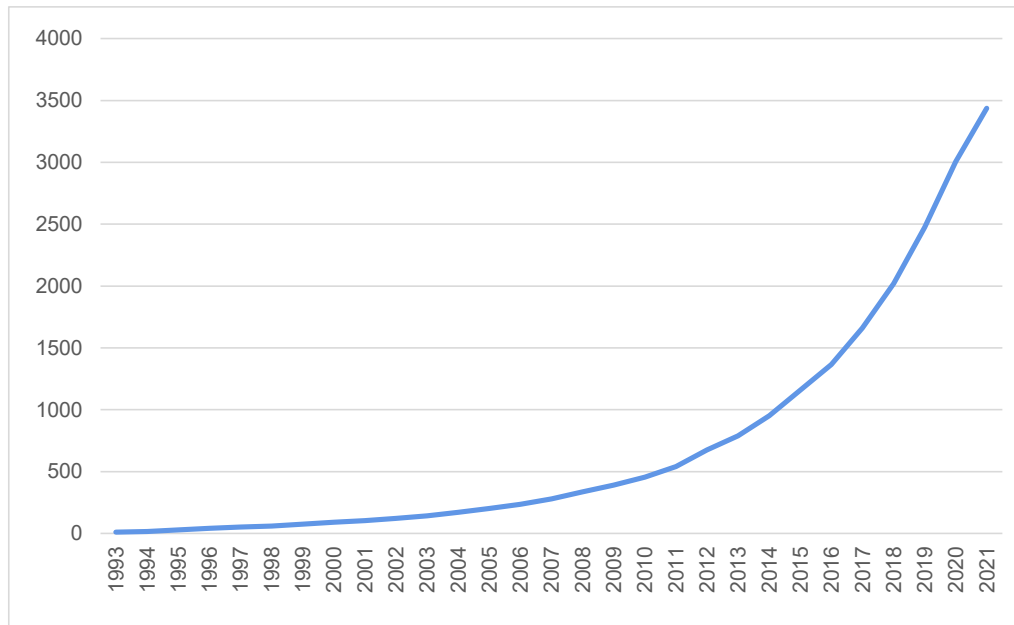
The word “**green logistics**” was used as search term, and the “**topic**” was used as the search route, the search year is from the 1993(the earliest year in this website) to 2021. Then, **3437 articles** were obtained.

Finally, the “**title**” and “**abstract**” of the document are combined as the text to be analyzed by R programming.

4 Experiment and results

4.1 Development stage division

After obtaining the required data, the **cumulative number of articles at each year** is calculated. A life cycle S-curve is drawn based on the results, as shown in Figure.



Based on the figure, we can divide the development of green logistics field into three stages.

- The **emerging stage** is 1993 -2003.
- The **slow growth stage** is 2004 -2014.
- The **rapid growth stage** is 2015 - now.

Fig 4. Technology life cycle S-curve of green logistics

4 Experiment and results

4.2 Topic analysis - The emerging stage

Table 3. Related words and meaning of **emerging stage (1993-2003)** topics

Topic	Words	Meaning
1-1	Risk; green; studi; logist; use; associ; model; growth; regress; intak; rate; speci; factor; data; develop	Elements of developing GL
1-2	Consumpt; environ; ratio; found; high; yield; portal; reduc; trend; treatment; adjust; describ; statist; measur; densiti	Measures to reduce environmental consumption
1-3	Function; behavior; chang; examin; indic; role; fruit; mean; improv; obtain; presenc; simul; requir; develop; flower	The Influencing factors of human behavior on GL
1-4	Relationship; purpo; produc; evalu; global; inform; product; anali; compar; less; data; leav; oper; base; effici	Relationship analysis between GL and other issues

4 Experiment and results

4.2 Topic analysis - The emerging stage

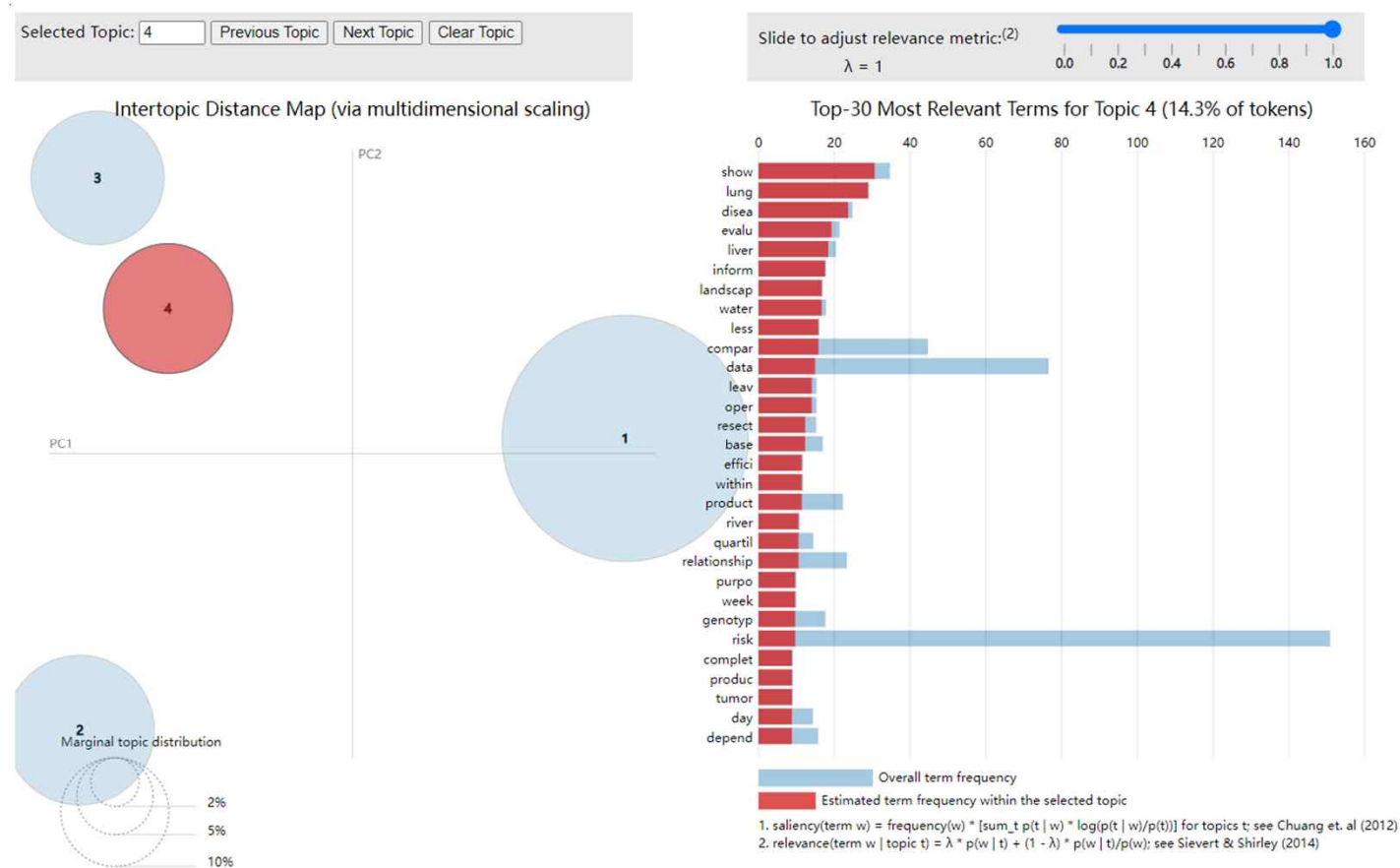


Fig 7. Visualization of LDA topic model in emerging stage (1993-2003)

4 Experiment and results

4.2 Topic analysis - The slow growth stage

There are 810 articles in web of science on the slow growth stage of green logistics.

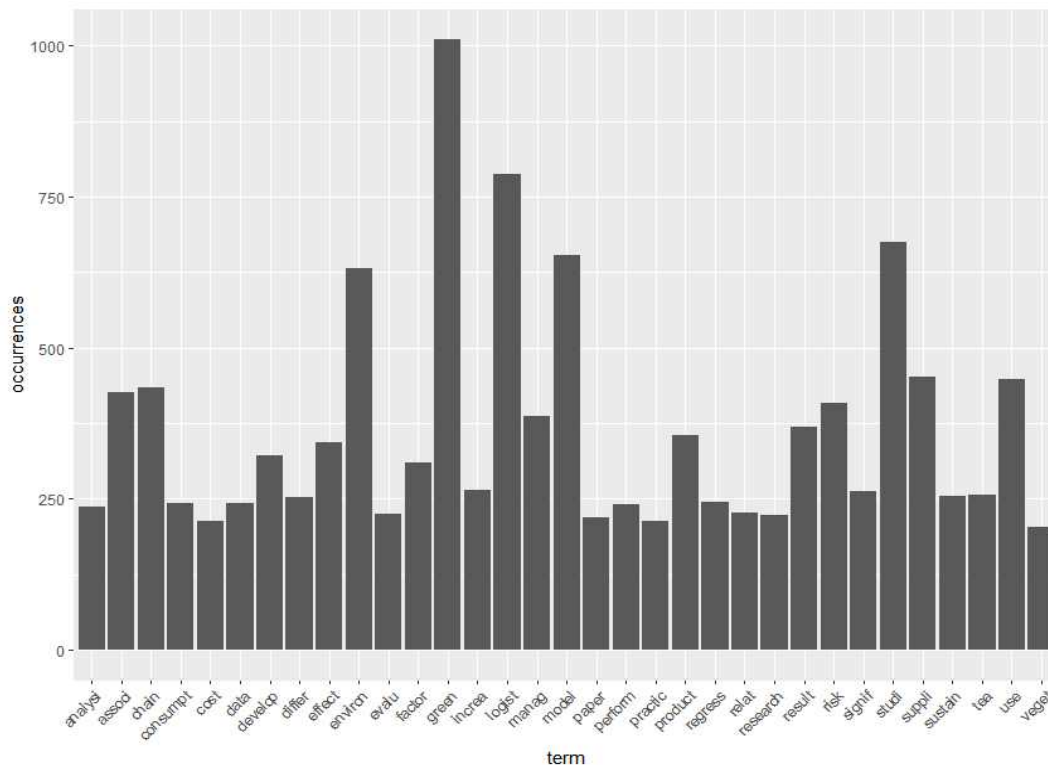


Fig 8. The histogram of most frequency in slow growth stage

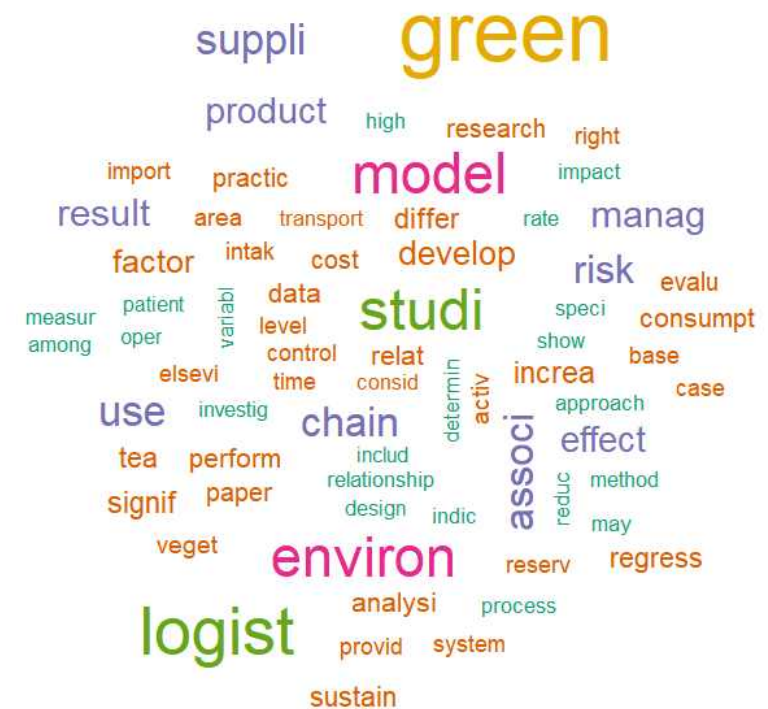


Fig 9. The wordcloud of most frequency in slow growth stage

4 Experiment and results

4.2 Topic analysis - The slow growth stage

Table 4. Related words and meaning of **slow growth stage (2004-2014)** topics

Topic	Words	Meaning
2-1	Green; logist; environ; suppli; chain; model; manag; product; perform; practic; paper; research; sustain; cost; transport	Research on Sustainable Development Management of GL
2-2	Associ; intak; speci; problem; impact; risk; Ltd; built; compani; econom; inver; express; approach; increa; among	Problems between logistic company, economy and environment
2-3	Studi; use; risk; associ; effect; model; factor; result; signif; vehicl; increa; regress; data; evalu; differ	Significant factors affecting logistics efficiency
2-4	Manufactur; consid; market; approach; deci; carbon; framework; literatur; vehicl; firm; analysi; busi; issu; general; method	The status quo of enterprise manufacturing and transportation
2-5	Closedloop; reserv; case; contain; gscm; life; port; solvDeci; reveal; valid; barrier; shipper; encrypt; issu	Types and cases of green supply chain

4 Experiment and results

4.2 Topic analysis - The slow growth stage

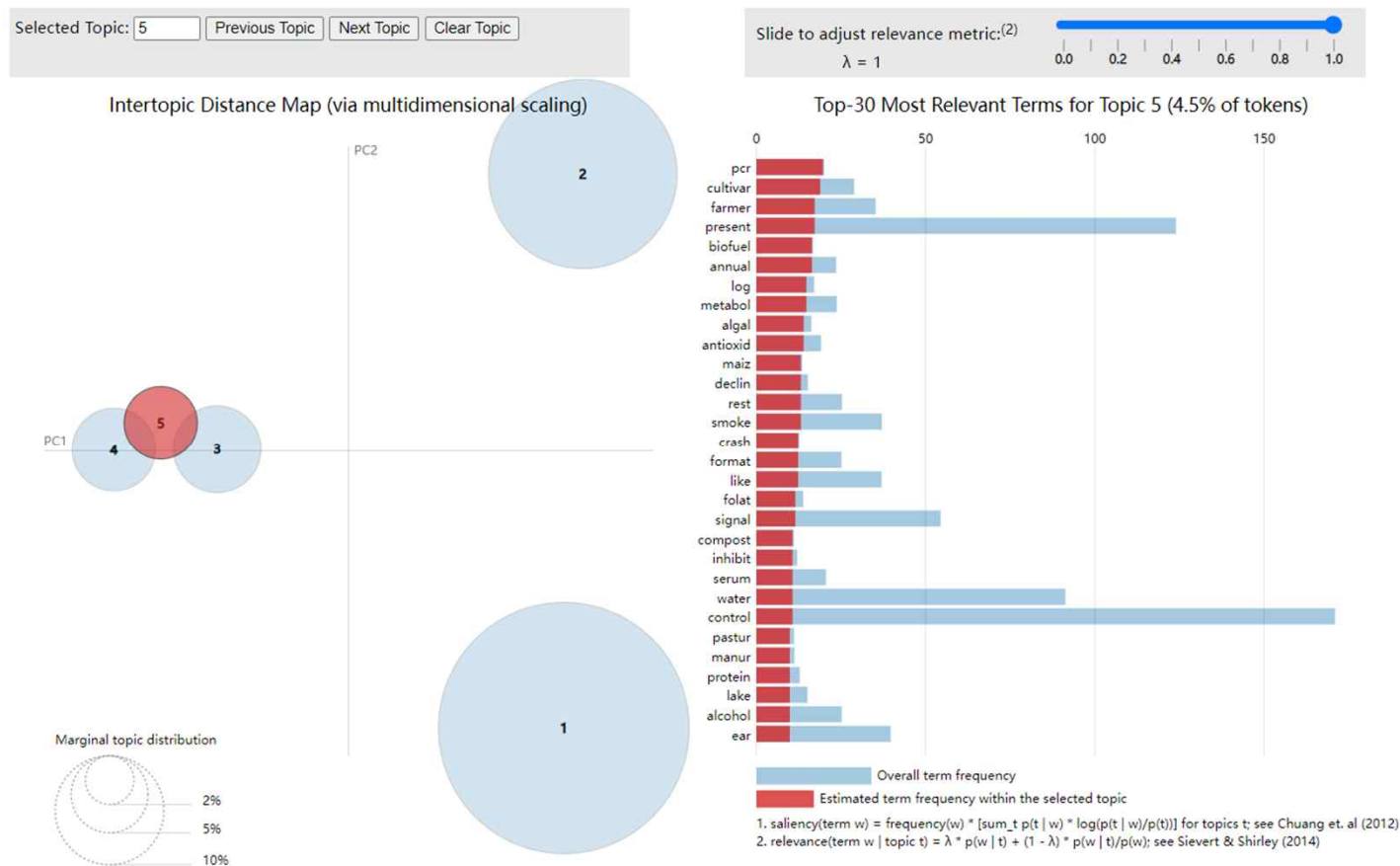


Fig 10. Visualization of LDA topic model in slow growth stage (2004-2014)

4 Experiment and results

4.2 Topic analysis - The rapid growth stage

Table 5. Related words and meaning of **rapid growth stage (2015-now)** topics

Topic	Words	Meaning
3-1	Model; environ; green; studi; chain; suppli; logist; sustain; develop; manag; product; result; use; research; perform	Model and management on sustainable development of Green Logistics
3-2	Associ; studi; risk; analysi; factor; use; regress; data; model; urban; differ; among; signif; predict; chang	Related issues and potential threats of GL
3-3	Effect; optim; evalu; impact; factor; vehicl; improv; carbon; increa; process; approach; oper; differ; time; found	Optimization analysis of low-carbon vehicle routing and time
3-4	Logist; green; practic; perform; enterpri; rever; servic; oper; econom; rout; develop; distribut; gscm; compani; packag	The performance of logistic company in GSCM
3-5	Factor; valu; imag; intersect; adopt; import; yield; gas; reduct; network; local; quantit; sscm; perceiv; integr	Green Supply Chain Network Analysis
3-6	Gscm; recycl; signif; product; driver; ban; gsc; plant; pretreat; indic; strategi; effici; select; polici; measur	Strategy and policy of GSCM
3-7	Green; gscm; space; object; associ; barrier; servic; mental; data; yield; network; solut; road; gsc; disrupt	The Dilemma of green logistics development

4 Experiment and results

4.2 Topic analysis - The rapid growth stage

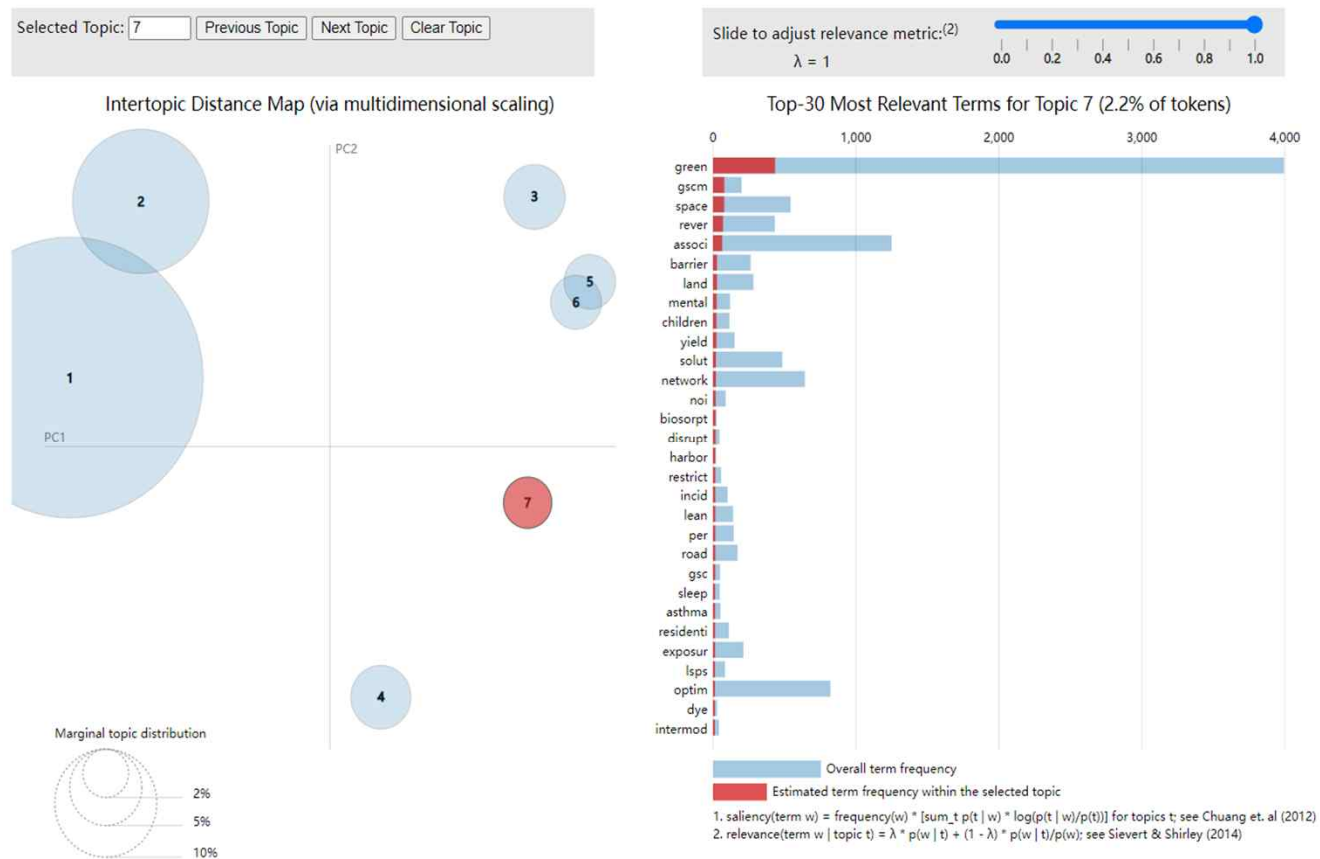


Fig 13. Visualization of LDA topic model in rapid growth stage (2015-now)

4 Experiment and results

4.3 Topic similarity calculation result and visualization

This study uses **cosine similarity** to calculate the **relationship between various topics**, and then to identify the evolutionary relationships between different technology topics in different technology stages. The topic similarity matrix can be seen in table 6 and table 7.

We set a **threshold value of 0.4** for determining the similarity between two topics. **If the similarity is greater or equal to 0.4, there is a link between two topics.** Based on the final results of all calculation , we can get the final roadmap in Fig 14.

4 Experiment and results

4.3 Topic similarity calculation result and visualization

Table 6. Similarity between **emerging stage** and **slow growth stage**

Topic	2-1	2-2	2-3	2-4	2-5
1-1	0.63	0.54	0.45571429	0.14857143	0.11857143
1-2	0.16571429	0.43857143	0.28142857	0.12428571	0.09714286
1-3	0.18661044	0.1182341	0.26780735	0.48404593	0.17122536
1-4	0.16547806	0.12268201	0.24964362	0.49557784	0.56847368

Table 7. Similarity between **slow growth stage** and **rapid growth stage**

Topic	3-1	3-2	3-3	3-4	3-5	3-6	3-7
2-1	0.6523278	0.4741782	0.13642683	0.27020241	0.40927392	0.43576457	0.14238723
2-2	0.40596478	0.30861385	0.12318151	0.09073047	0.11258525	0.10728712	0.41589658
2-3	0.51855442	0.33709568	0.49536852	0.16357975	0.14172496	0.14503629	0.14702309
2-4	0.16622881	0.42477226	0.10000219	0.4880814	0.0960286	0.47814741	0.09271726
2-5	0.25166115	0.18675906	0.40000219	0.41390978	0.20132673	0.10530032	0.11589658

4 Experiment and results

4.3 Topic similarity calculation result and visualization

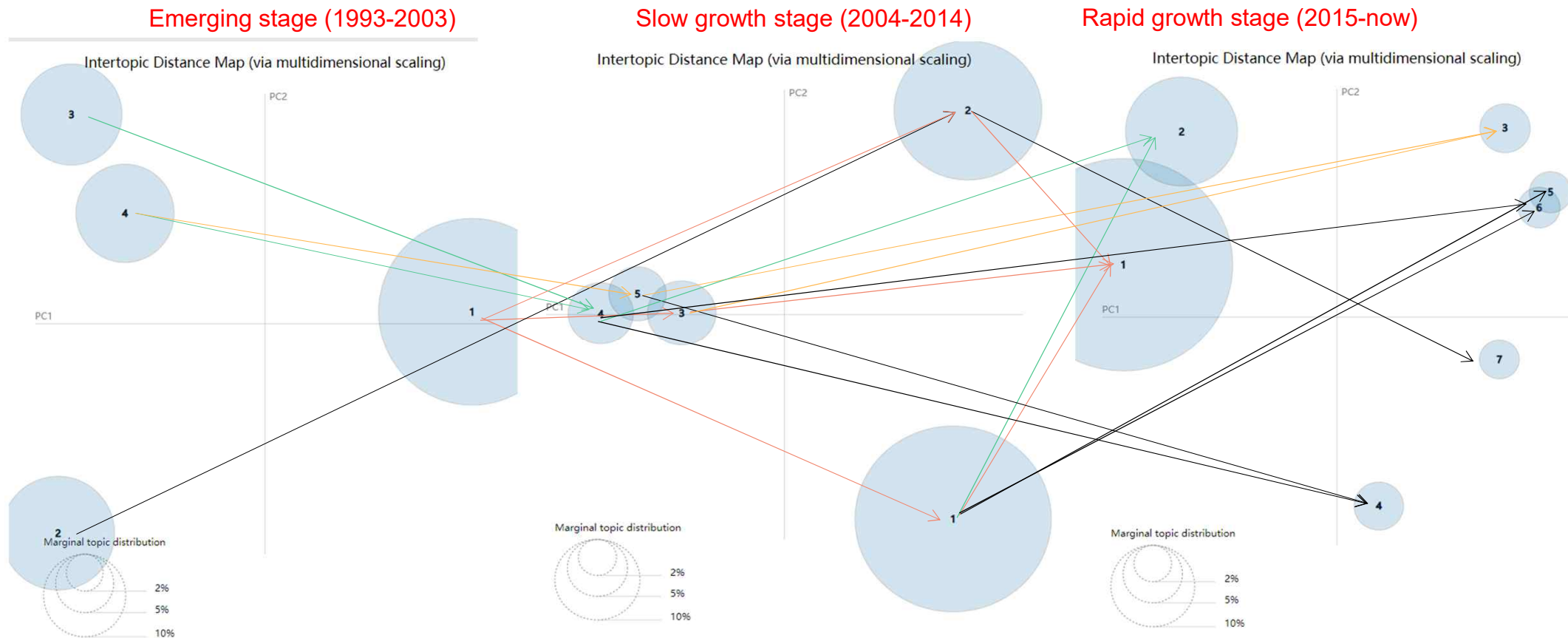


Fig 14. LDA-based Roadmap of Green Logistics

4 Experiment and results

4.3 Topic similarity calculation result and visualization

- **Topic 3-1 (model and management on sustainable development of green logistics)** evolved from topic 2-1 (research on Sustainable Development Management of GL), topic 2-2 (problems between logistic company, economy and environment) and topic 2-3 (significant factors affecting logistics efficiency) in stage 2 and topic 1-1 (elements of developing green logistics) in stage 1.
- **Topic 3-2 (related issues and potential threats of GL)** evolved from topic 2-1 (research on Sustainable Development Management of GL) and topic 2-4 (the status quo of enterprise manufacturing and transportation) in stage 2, topic 1-1 (elements of developing GL) , topic 1-3 (the Influencing factors of human behavior on GL) and topic 1-4 (relationship analysis between GL and other issues) in stage 1.
- **Topic 3-3 (Optimization analysis of low-carbon vehicle routing and time)** evolved from topic 2-3 (Significant factors affecting logistics efficiency) and topic 2-5 (Types and cases of green supply chain) in stage 2, topic 1-1 (Elements of developing GL) and topic 1-4 (Relationship analysis between GL and other issues) in stage 1.

4 Experiment and results

4.3 Topic similarity calculation result and visualization

- **Topic 3-4 (The performance of logistic company in GSCM)** evolved from topic 2-4 (The status quo of enterprise manufacturing and transportation) and topic 2-5 (Types and cases of green supply chain) in stage 2, and topic 1-3 (The influencing factors of human behavior on GL) and topic 1-4 (Relationship analysis between GL and other issues) in stage 1.
- **Topic 3-5 (Green Supply Chain Network Analysis)** evolved from topic 2-1 (Research on sustainable development management of GL) in stage 2 and topic 1-1 (Elements of developing green logistics) in stage 1.
- **Topic 3-6 (Strategy and policy of GSCM)** evolved from topic 2-1 (Research on sustainable development management of GL) and topic 2-4 (The status quo of enterprise manufacturing and transportation) in stage 2, topic 1-1 (), topic 1-3 (The influencing factors of human behavior on GL) and topic 1-4 (Relationship analysis between GL and other issues) in stage 1.
- **Topic 3-7 (The Dilemma of green logistics development)** evolved from topic 2-2 (Problems between logistic company, economy and environment) in stage 2, topic 1-1 (Elements of developing green logistics) and topic 1-2 (Measures to reduce environmental consumption) in stage 1.

5 Conclusions

This article divides the development of green logistics into three stages, which are emerging stage (1993-2003), slow growth stage (2004-2014) and rapid growth stage (2015-now).

Based on LDA topic model and text similarity, this study got the main topics in each three stage. The main contents of the current development of green logistics include “model and management on sustainable development of green logistics”, “related issues and potential threats of green logistics”, and “optimization analysis of low-carbon vehicle routing and time”, which evolved from topics in emerging stage and slow growth stage.

“The performance of logistic company in GSCM”, “Green supply chain network analysis”, “strategy and policy of GSCM” and “the Dilemma of green logistics development” are important points that people will continue to study and pay attention to in the future.

6 Reference

- Lee S Y, Klassen R D. Drivers and enablers that foster environmental management capabilities in small-and medium-sized suppliers in supply chains[J]. Production and Operations management, 2008, 17(6): 573-586.
- Marvi, Mohammad Hossein, et al. "Knowledge foundation in green purchase behaviour: Multidimensional scaling method." Cogent Business & Management 7.1 (2020): 1773676.
- Freis J, Vohlidka P, Günthner W A. Low-carbon warehousing: Examining impacts of building and intra-logistics design options on energy demand and the CO2 emissions of logistics centers[J]. Sustainability, 2016, 8(5): 448.
- Straka M, Spirkova D, Filla M. IMPROVED EFFICIENCY OF MANUFACTURING LOGISTICS BY USING COMPUTER SIMULATION[J]. International Journal of Simulation Modelling (IJSIMM), 2021, 20(3).
- Fang X, Mu B, Yu C. Strengthening green transportation and implementing sustainable urban environmental development in the new period[J]. International Journal of Environmental Technology and Management, 2020, 23(2-4): 149-161.
- Pani A, Mishra S, Golias M, et al. Evaluating public acceptance of autonomous delivery robots during COVID-19 pandemic[J]. Transportation research part D: transport and environment, 2020, 89: 102600.
- Sellitto M A, de Almeida F A. Strategies for value recovery from industrial waste: case studies of six industries from Brazil[J]. Benchmarking: An International Journal, 2019.
- Vayansky I, Kumar S A P. A review of topic modeling methods[J]. Information Systems, 2020, 94: 101582.
- Zhang H, Daim T, Zhang Y P. Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain[J]. Technological Forecasting and Social Change, 2021, 167: 120729.
- Du Y J, Yi Y T, Li X Y, et al. Extracting and tracking hot topics of micro-blogs based on improved Latent Dirichlet Allocation[J]. Engineering Applications of Artificial Intelligence, 2020, 87: 103279.
- Bastani K, Namavari H, Shaffer J. Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints[J]. Expert Systems with Applications, 2019, 127: 256-271.
- Roque C, Cardoso J L, Connell T, et al. Topic analysis of road safety inspections using latent Dirichlet allocation: A case study of roadside safety in Irish main roads[J]. Accident Analysis & Prevention, 2019, 131: 336-349.
- Ding S, Li Z, Liu X, et al. Diabetic complication prediction using a similarity-enhanced latent Dirichlet allocation model[J]. Information Sciences, 2019, 499: 12-24.

Thank You