

A dynamic network analytic method for identifying emerging technologies

Lu Huang*, Yihe Zhu, Xingxing Ni

School of Management and Economics, Beijing Institute of Technology

Yi Zhang

Centre for Artificial Intelligence, University of Technology Sydney

Background

- Emerging technologies, such as artificial intelligence and quantum computing, create **opportunities** and **challenges** to science policy and entrepreneurships.
- **Identifying emerging technologies** at an early stage would provide benefits for governments and companies to take preemptive actions for forthcoming competitions.

What is an emerging technology^[1]?

- (1) Radical **novelty** ;
- (2) Relative **fast growth** ;
- (3) A certain degree of **coherence** with the passage of time ;
- (4) **Prominent impact** to the technical field ;

[1] Rotolo D, Hicks D, Martin B R. What is an emerging technology?[]]. Research Policy, 2015, 44(10):1827-1843.

Previous studies

Citations analysis	Cho et al. represented the technical field with IPC classification number, established 128 reference networks in Taiwan, and considered the Structural Holes in the network as an emerging technology.
Co-word analysis	Lee got the technical topic network through hierarchical clustering method in keyword co-occurrence network , and then got the low degree, closeness and betweenness as the new technology through network analysis.
Hybrid (citations+co-word)	Glkalnzal et al. mixed direct citation relation and keyword coupling relation to cluster, and holds that emerging technology is the research topic of the changing field and emerging technology topic.



Research Issues

- (1) **Characteristics** of emerging technologies are usually **unobvious** – they keep changing
- (2) Previous studies emphasize **post-event identifying nature** ^[2] – i.e., historical profiling
- (3) Understandings on emerging technologies vary between researchers and disciplines ^[3].



Our Tasks

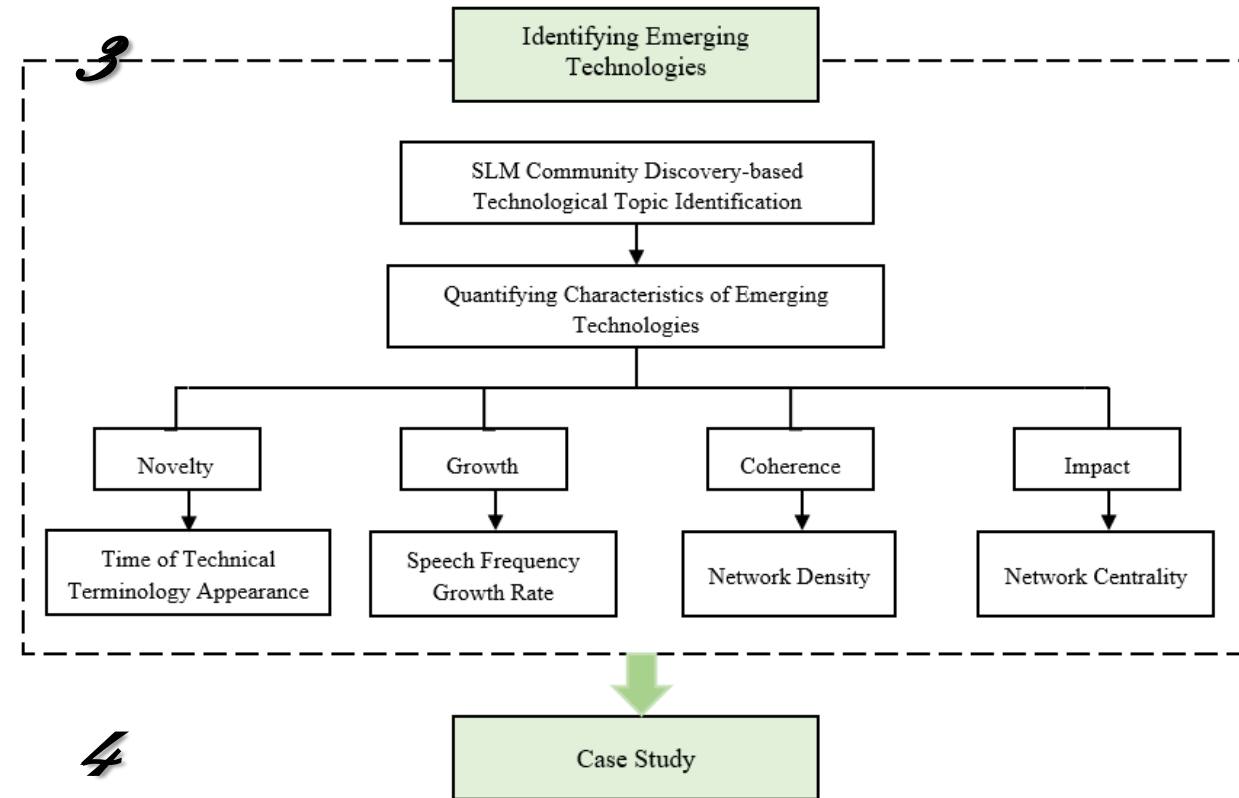
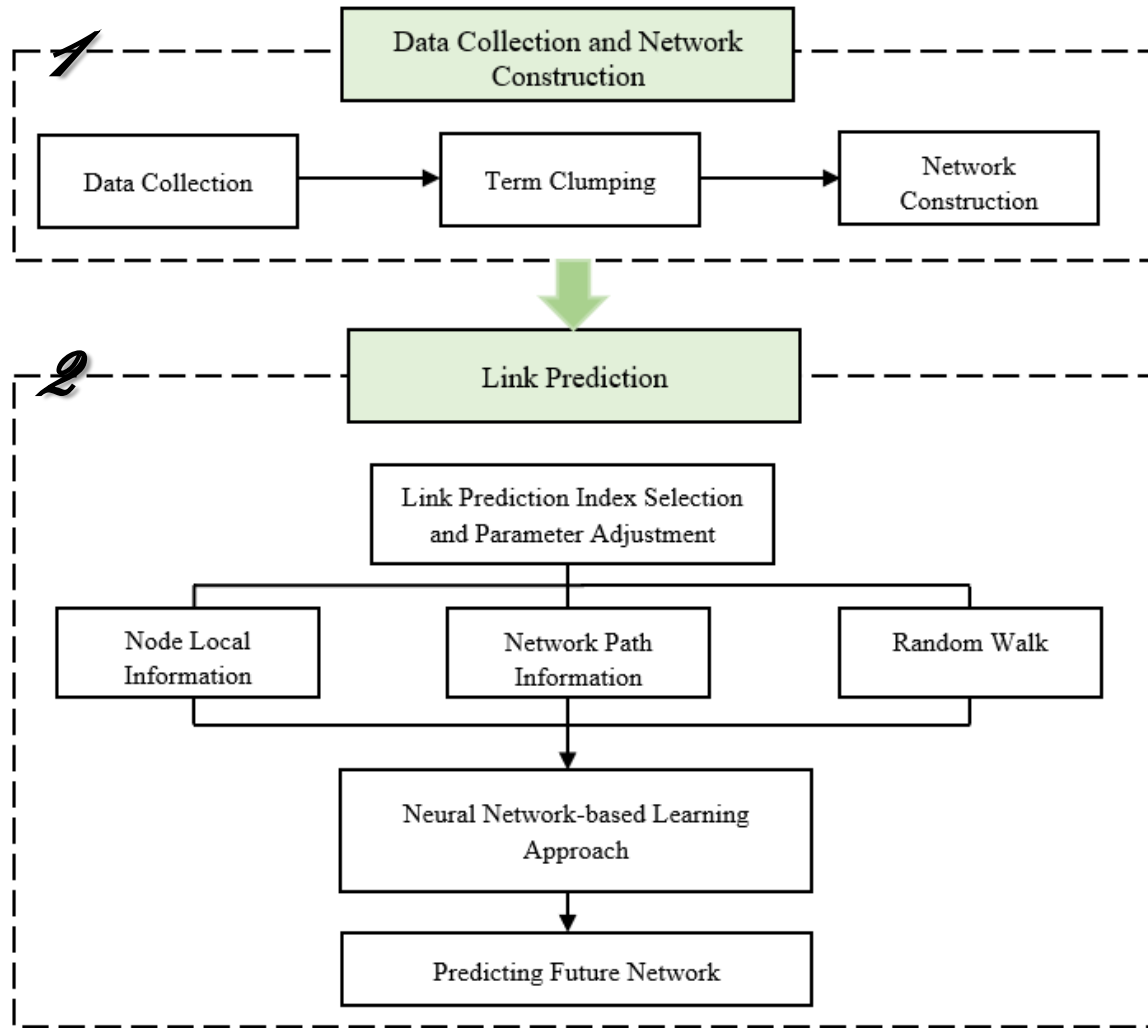
- (1) Quantifying the characteristics of emerging technologies via bibliometric indicators
- (2) Conducting pre-event identifying nature via link prediction and machine learning techniques

[2] Cozzens S , Gatchair S , Kang J , et al. Emerging technologies: quantitative identification and measurement[J]. Technology Analysis & Strategic Management, 2010, 22(3):361-376.

[3] Qi W . A bibliometric model for identifying emerging research topics[J]. Journal of the Association for Information Science & Technology, 2017.



Research Framework



Method- Data Collection and Network Construction

1. Collecting Data from the WoS;
2. Using VantagePoint to remove noise and consolidate technological synonyms by using term clumping;
3. Constructing dynamic networks – i.e., a series of weighted co-word network based on time periods.

Method- Link Prediction

Task – Predicting a near-future network based on link prediction and machine learning approaches

- **Link Prediction Index Selection and Parameter Adjustment**

- **Common Neighbors (CN)**
- Surrounding Nodes

$$S_{xy}^{CN} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} w_{xz}^{\alpha} + w_{yz}^{\alpha} \quad (-1 \leq \alpha \leq 1)$$

- **Local Path (LP)**
- Pathways

$$S_{xy}^{LP} = \sum_{l_{xy}^n \in L(x,y)} l_{xy}^2 + \alpha l_{xy}^3$$

$$l_{xy}^n = \left(\sum_n w_{ij}^{-1} \right)^{-1}$$

- **SimRank (SR)**
- Dynamic Status

$$S_{xy}^{SR} = \alpha \frac{\sum_{v_z \in \Gamma(x)} \sum_{v_{z'} \in \Gamma(y)} S_{zz'}^{SR}}{k_x k_y}$$

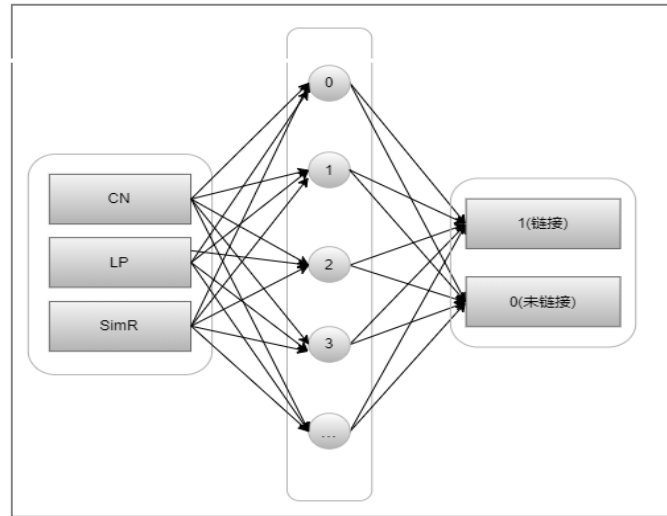
Method- Link Prediction

- **Neural Network-based Learning Approach for Parameter Training**

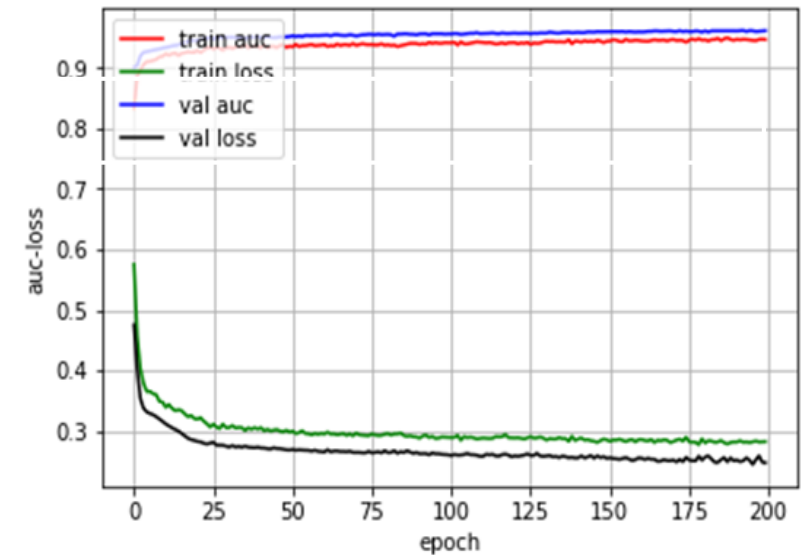
- The task of this model is to train the hyper-parameters in calculating the three indices.
- Considering the number of features and this two-class problem, a three-layer neural network is used, with 3 and 2 neurons in the input layer and the output layer, respectively. The number of neurons in the hidden layer was determined by trial-and-error method. The loss function is cross-entropy.
- The Area Under Curve (AUC) value is used for validation.

- **Predicting Future Network**

- Predicting the probability of a new edge appearing between two nodes.



Structure of BP Neural Network



Sketch of training process

Method-Identifying Emerging Technologies

- **Smart Local Moving (SLM) algorithm-based Community Discovery**

- The SLM algorithm is the function integrated in the VOSViewer^[5]

- **Four Characteristics of Emerging Technologies**

- **Novelty** – How new this technological topic is;
- **Growth** – How fast the frequency of terms associated with this technological topic;
- **Coherence** – How high the density of this technological topic is

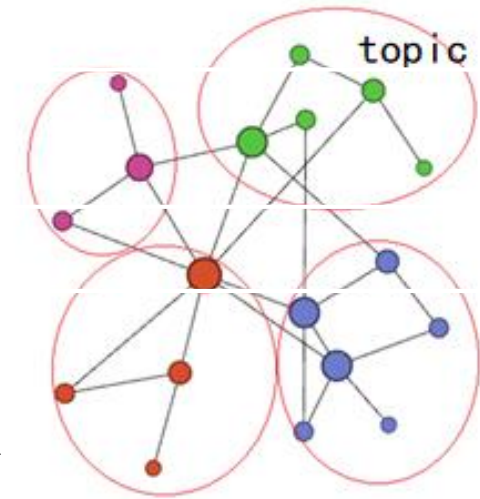
$D(i)$ = The number of Edges in this topic / The total number of Edges this topic could have

$D(e)$ = The number of Edges between this topic and other topics / The total number of Edges this topic could have with other topics

$\text{Coherence} = D(i) / D(e)$

- **Impact** – How soon the importance of this topic appears

A PageRank Algorithm for calculating the centrality of the topic with a time-biased weighting function



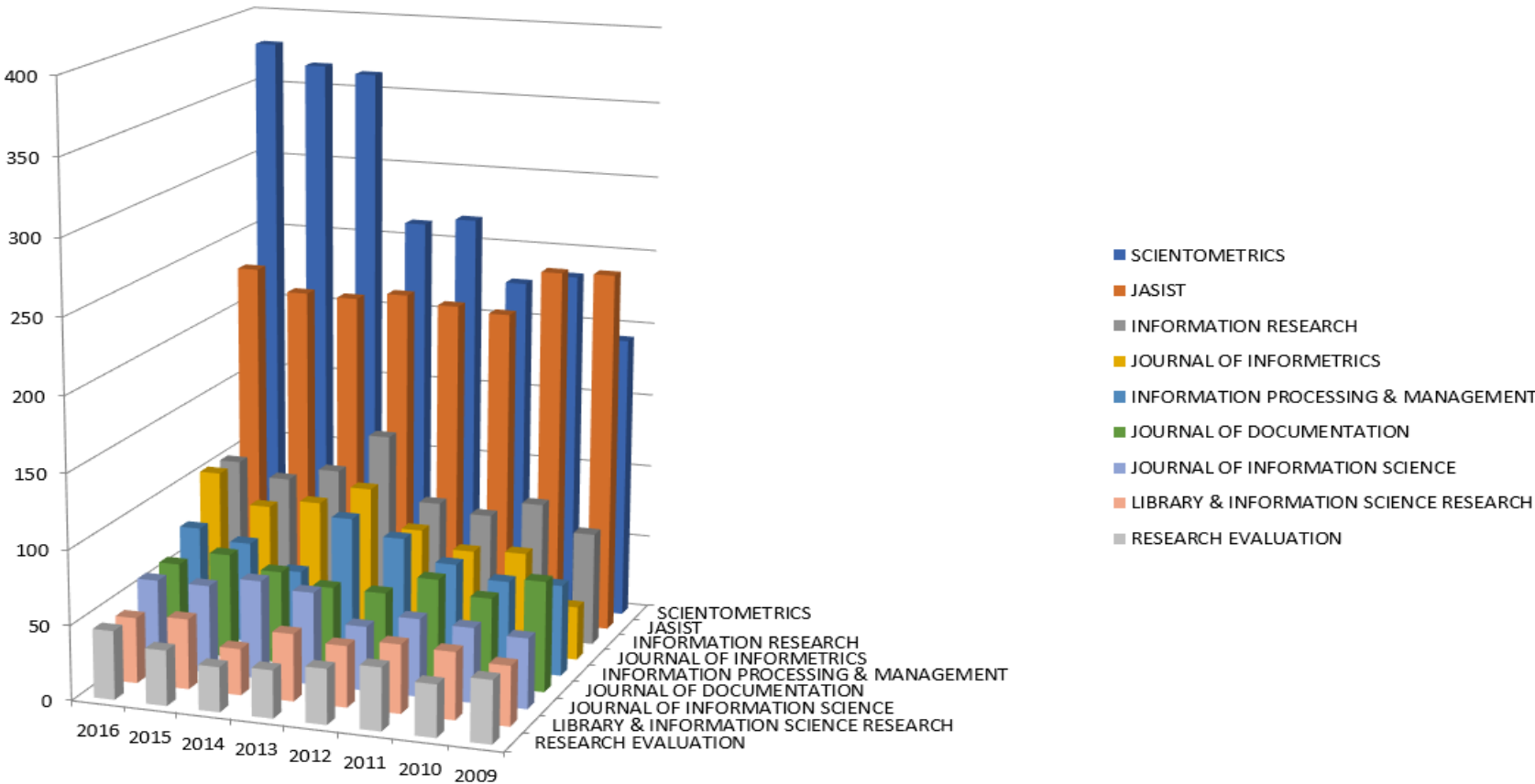
Case Study

Data source: core journals in the field of Information Science

Journal

- Scientometrics
- JASIS/JASIST
- Information Research an International Electronic Journal
- Journal of Informetrics
- Information Processing and Management
- Journal of Documentation
- Journal of Information Science
- Library and Information Science Research
- Research Evaluation
- ARIST

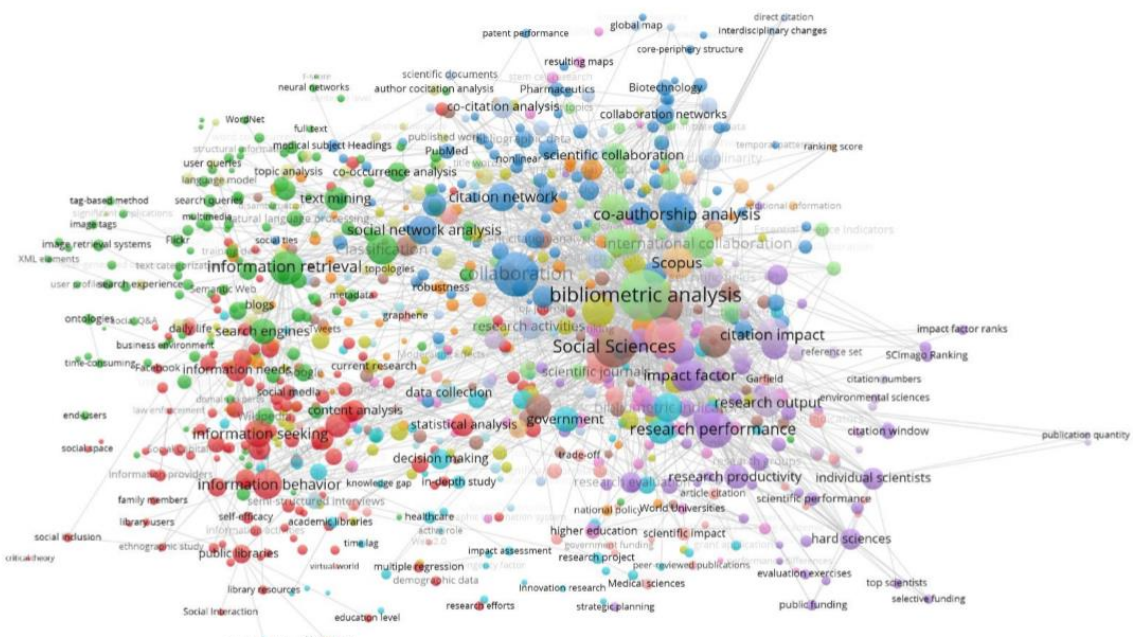
Journal data for 2009-2016



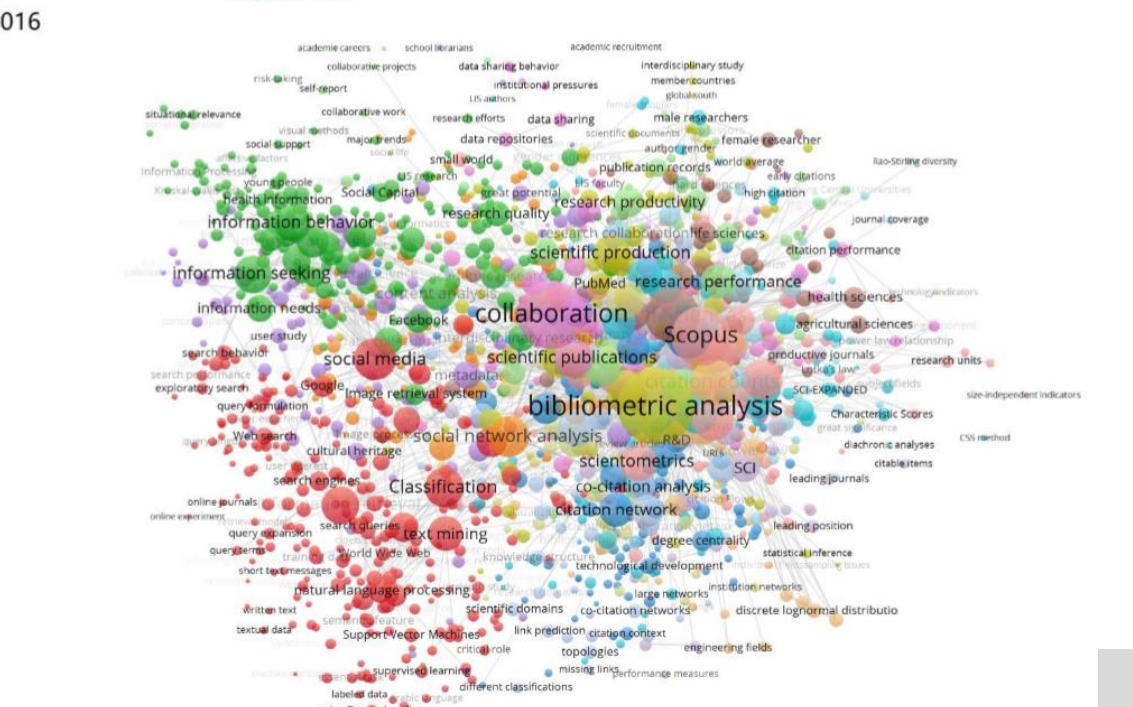
Case Study

Steps	Term Clumping	Number of words
1	The Natural Language Processing module of VP software integrates the original words.	152468
2	Remove numbers, words at the beginning and end of numbers. For example, "step 1" and "one way" and "several indices"	135157
3	Remove suspensions, prepositions and conjunctions: Stop words in commonly used English words, words and phrases commonly used in scientific and technical documents	128110
4	Remove generic words and common words in scientific and technological literature: For example, "research framework" and "novel method"	119641
5	Based on the synonyms of expert knowledge integration, For example, "co-word analysis" and "word co-occurrence analysis"	115601
6	Semantic Fuzzy Matching: Combining words with the same stem. For example, "citation count" and "association counting", "information seeking" and "seeking information"	102502
7	Delete words with frequency less than 3	5838
8	Words unrelated to research, such as "well-known measures" and "traditional methods", are artificially removed.	4640

2011-2012

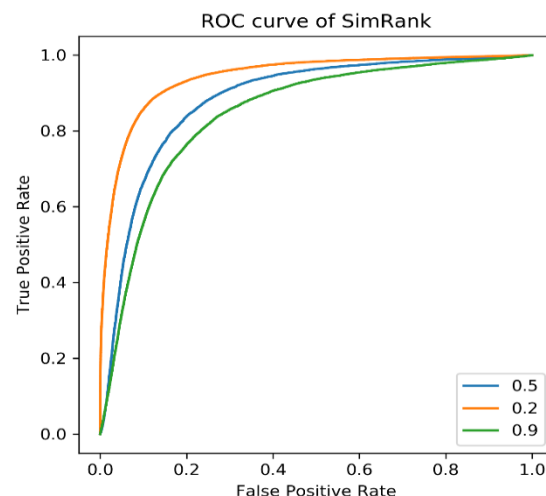
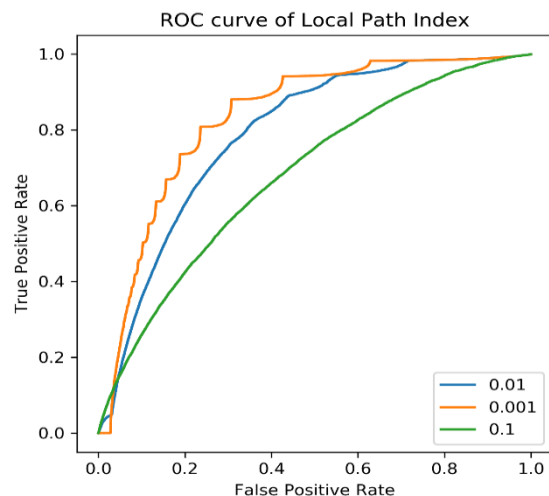
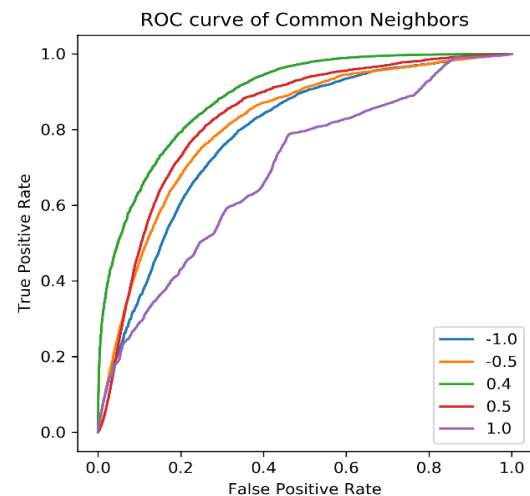


2015-2016

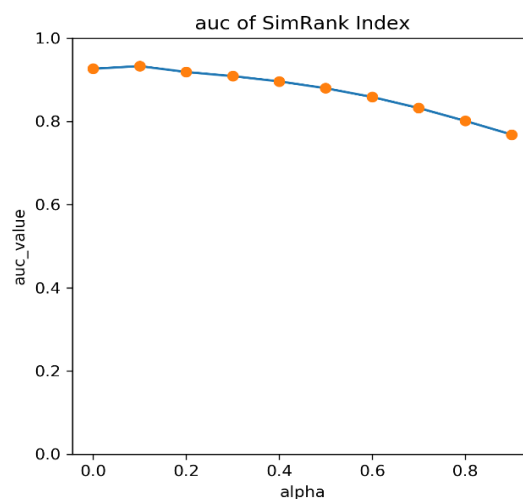
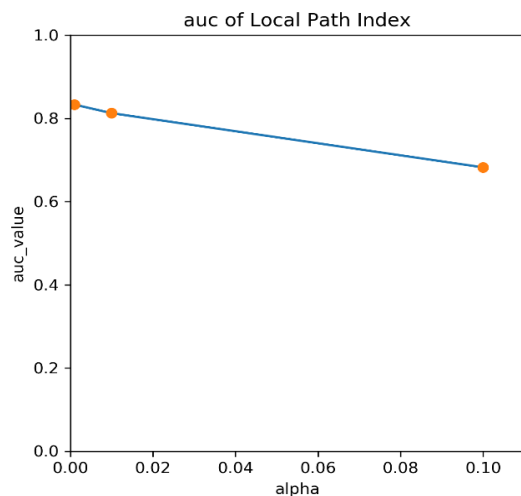
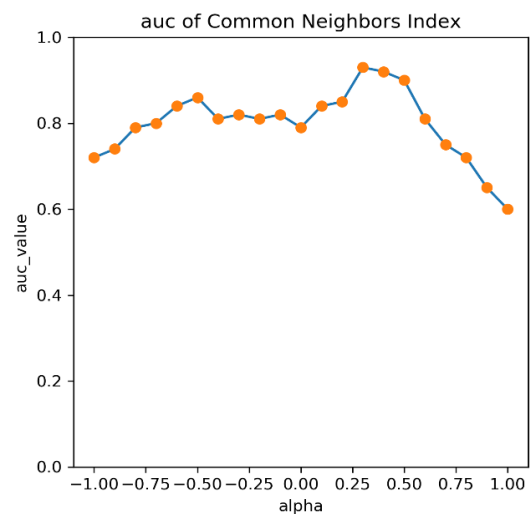


Case Study

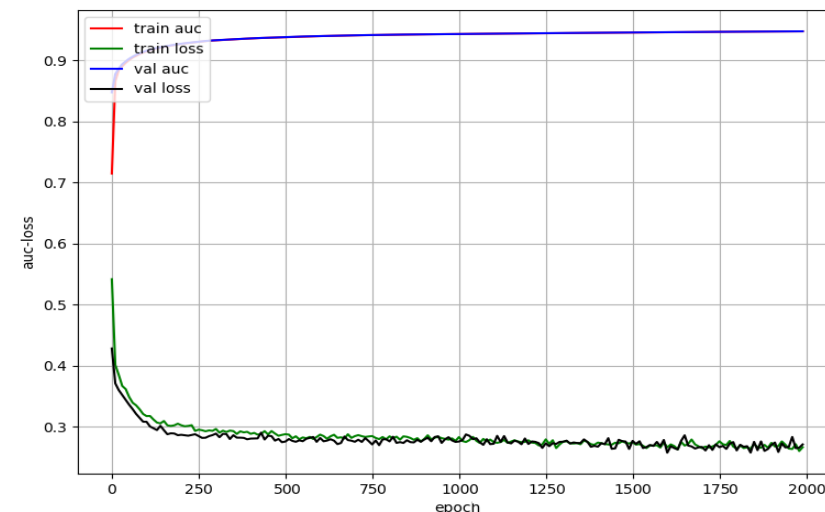
Link Prediction Index Parameter Training and Future Network Prediction



The link prediction model is applied to the network from 2015 to 2016. In the prediction network, 3525 new edges are predicted and a total of 35590 edges are generated.



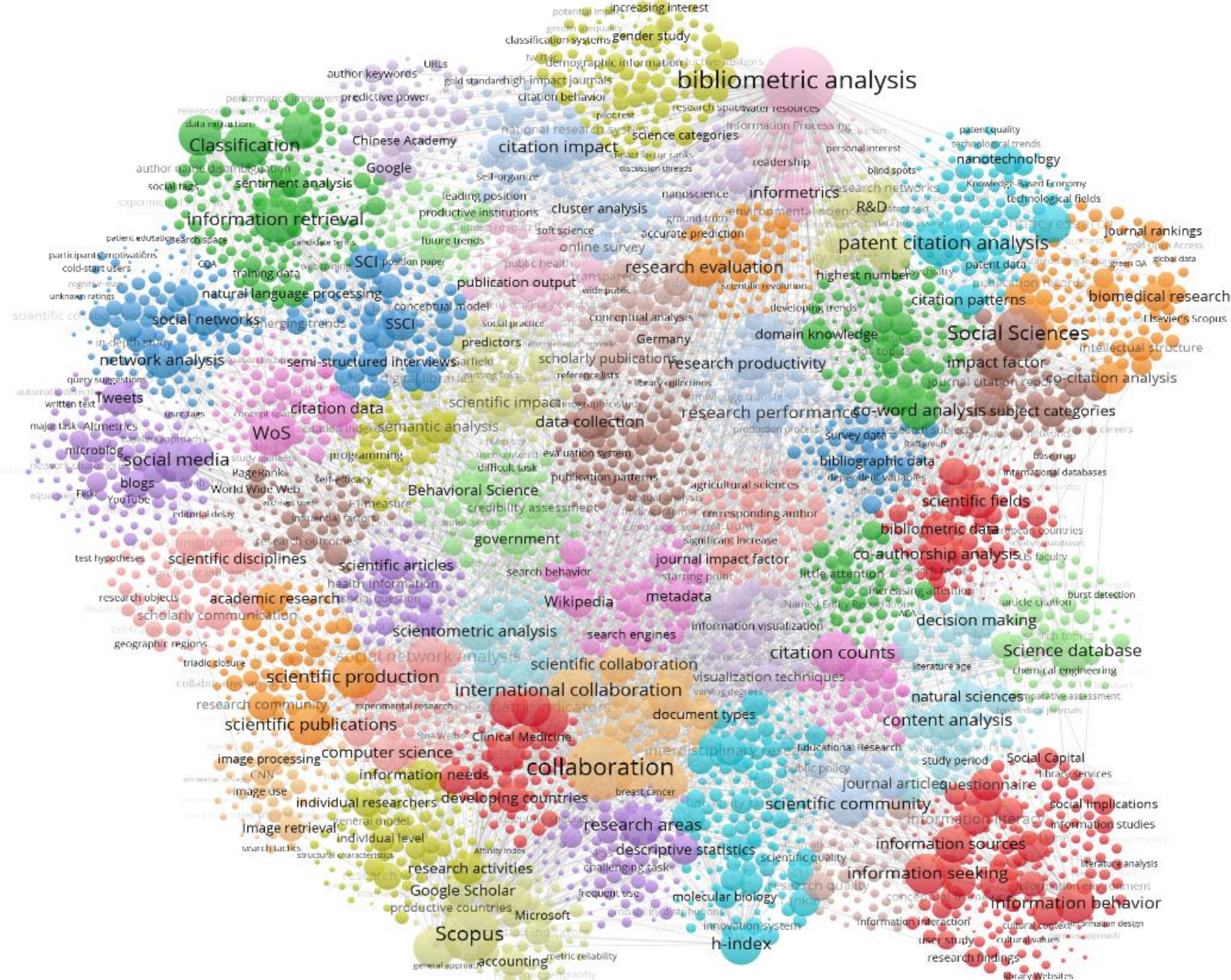
The alpha values of CN, LP and SimR are set to 0.4, 0.001 and 0.1.
The highest AUC values were 0.91, 0.82 and 0.92 respectively.



Training of BP Neural Network Prediction Model

Case Study

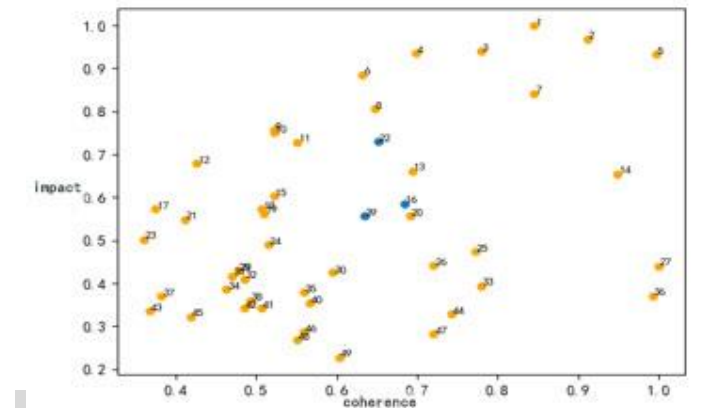
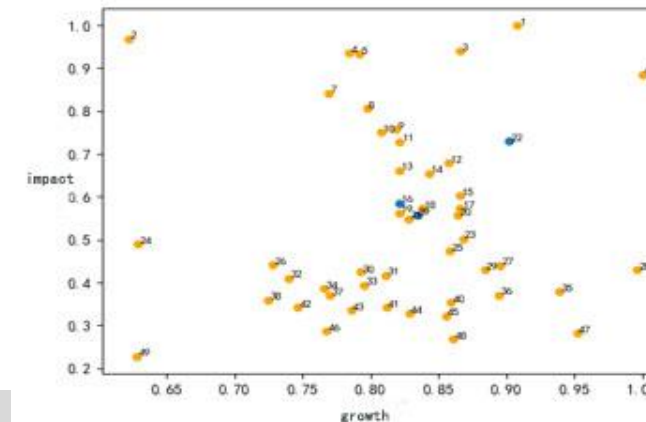
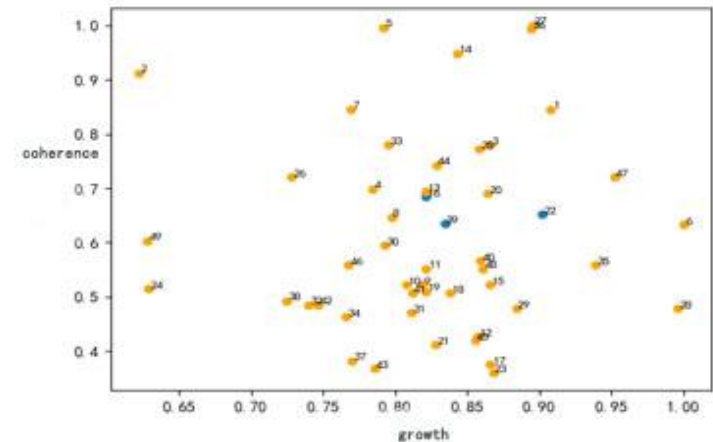
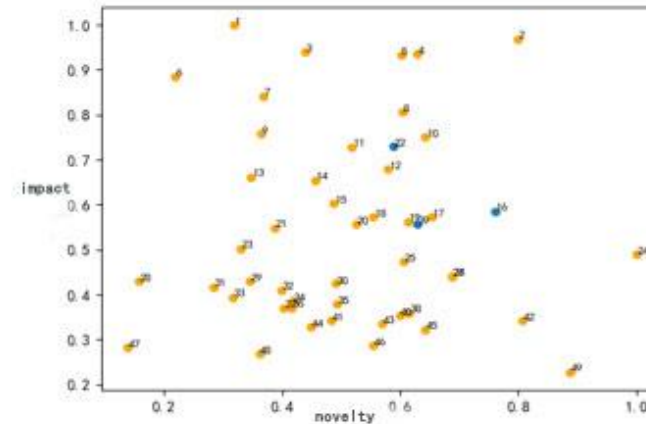
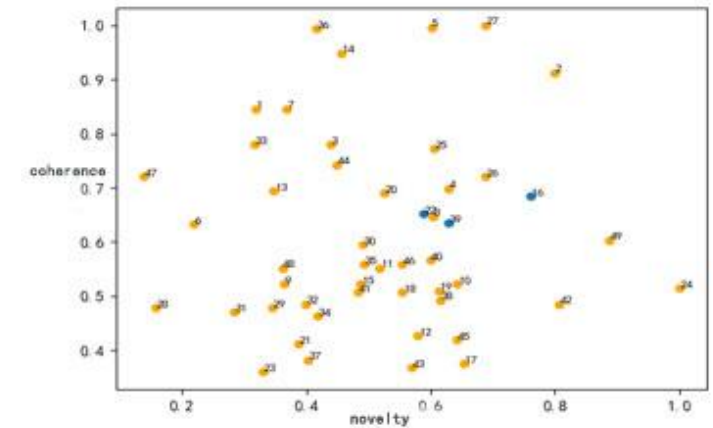
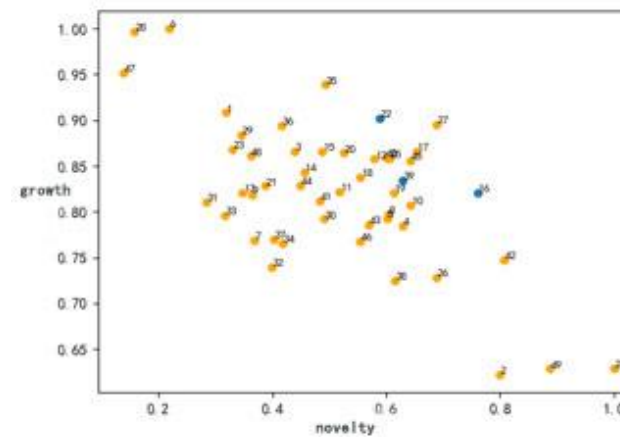
49 Technological Topics in Information Science with Predicted Links



Case Study

Emerging Technology Identification

- Each 2D map indicates two characteristics of emerging technologies
- Several emerging topics are observed:
 - Image Retrieve System (#16)
 - Altmetrics (#22)
 - Semantic Analysis (#39)



- **Highlights**

- The use of link prediction and neural networks provides a solution of identifying emerging technologies in a pre-event identifying nature – i.e., prediction, with limited expert involvement.
- The quantification of the four characteristics of emerging technologies provides a solution of characterizing emerging technologies with bibliometric indicators.

- **Future Directions**

- Validating the performance of the prediction through the records published in 2017 and after;
- Comparing the performance of the prediction with other network analytic baselines;
- Extending the case study from social science to natural science, and from publications to patents.

Thank you