Scientometrics (2022) 127:6637–6660
https://doi.org/10.1007/s11192-022-04330-9

Network dynamics in university-industry collaboration: a collaboration-knowledge dual-layer network perspective

Hongshu Chen[1] · Xinna Song[1] · Qianqian Jin[1] · Ximeng Wang[2]

Received: 15 August 2021 / Accepted: 23 February 2022 / Published online: 19 March 2022

Abstract

Collaborations between universities and firms provide a key pathway for innovation. In the big data era, however, the interactions between these two communities are being reshaped by information of much higher complexity and knowledge exchanges with more volume and pace. With this research, we put forward a methodology for comprehensively measuring both actor collaboration and produced knowledge in shaping network dynamics of university-industry collaboration. Using dual-layer networks consisting of organizations and topics, we mapped the longitudinal correlations between partnerships and knowledge in terms of both co-applications of patents and semantics. Network structures, individual characteristics, and knowledge proximity indicators were used to depict the longitudinal networks and then model the network dynamics. Further, a stochastic actor-oriented model was used to provide insights into the factors contributing to the network's evolution. A case study on university-industry collaborations in the information and communications technology sector demonstrates the feasibility of the methodology. The result of this study can be used for future research into the mechanisms that underpin university-industry collaborations and opportunity discovery.

Keywords Collaboration network · Knowledge network · University-industry collaboration · Network dynamics · Topic modeling · Doc2vec · Stochastic actor-oriented models

Introduction

University-industry collaborations are one of the most effective ways of boosting innovation capacity of organizations, countries or higher-level collectives (Nsanzumuhire & Groot, 2020). In the big data era, however, the interactions and knowledge exchanges between universities and industry have become more complex. As such, more and more studies are examining the dynamics of this vibrant landscape of co-innovation via a network perspective (Fischer et al., 2019; Mao et al., 2020; Chang, 2017). Their goal, and

✉ Hongshu Chen
Hongshu.Chen@bit.edu.cn

[1] School of Management and Economics, Beijing Institute of Technology, Beijing, China

[2] Cyber Finance Department, Postal Savings Bank of China, Beijing, China

ours, is to re-recognize and understand the patterns and dynamics of university-industry collaboration networks–to find the mechanisms of how these relationships evolve that yields theoretical and practical significance.

University-industry connections involve complex interactions of knowledge and resources. The innovations of the two communities are doubly embedded in the social networks of the participants and their knowledge networks (Guan & Liu, 2016; Wang & Hsu, 2014). The nodes and links in both collaborative networks and knowledge networks reveal the characteristics, content, and dynamics of the interactions between industry and academia. Many researchers have modeled the actor layer, i.e., individuals, teams, institutions, countries, and so on, with promising results. Commonly, complex network analysis is used via the co-authorship of research articles, co-applications of patents, co-funded grants, or contracts covering formal research cooperation (Guan & Zhao, 2013; Huang et al., 2015; Magazinik et al., 2019). Quantitative methods based on network analysis theory have also been used with some developing novel techniques to explore how these collaborations form and evolve (Akhtar et al., 2019; Yan & Guan, 2018; Zhang et al., 2016). In terms of the knowledge layer, however, the elements and relations formed by the content of these collaborations has only been examined in more recent years (Fischer et al., 2019; Guan et al., 2017; Hellsten & Leydesdorff, 2020). Moreover, most research models these two layers in separate strands, which only provides a partial view of the landscape.

Thus, we are faced with a new challenge of modeling both actors and knowledge in the same network. This requires extending the traditional single-layer networks into multilayer networks since one-mode networks cannot reflect the richness of the different kinds of interactions in real-world collaborations (Zhang & Ye, 2020). To examine the knowledge integration embedded in these relationships, researchers often map the content of joint publications or joint patents into knowledge networks. These content-based networks are mainly created via co-occurrences in international patent classification (IPC) codes (Chang, 2017; Yan & Guan, 2018), research article keywords (Guan et al., 2017), hashtags (Hellsten & Leydesdorff, 2020), and other straightforward tags that provide brief representations of the content. Yet, with these concise representations, it can be difficult to deliver the context of the textual data or provide enough information for further any semantic analysis. This is particularly so when the study relies on subjective or individual taxonomies (Roth & Cointet, 2010). To go a step further, constructing enhanced multilayer networks that portray richer knowledge requires a methodology steeped in topic extraction and semantic representation. Despite great progress, there is still much room to further improve today's methodologies. Word and document embedding algorithms are showing a very promising ability to map words and documents from large-scale text data into numeric vectors (Chen et al., 2022; Le & Mikolov, 2014; Mikolov et al., 2013; Zhang et al., 2018). These vectors can be used to replace traditional keyword representations in scientific text mining, thus holds great potential for topic extraction. Yet how to construct multiplex networks with these algorithms to yield rich insights of network dynamics urges deeper investigations for mechanism study of university-industry collaborations (Teng et al., 2021).

To fill the gap of comprehensively measuring both actors and knowledge in shaping university-industry collaboration network dynamics, we construct a series of dual-layer networks comprising both the collaborative entities and the knowledge elements that form the basis of their exchange. The proposed methodology integrates the theory and techniques of bibliometrics, scientific text mining, complex network analysis, and stochastic actor-oriented models (SAOMs) to comprehensively investigate the evolution of university-industry collaborations and their knowledge interactions from a dynamic perspective. The influential factors underpinning these dynamics are investigated in three respects:

individual characteristics, network structures, and proximity. Finally, a case study on the information and communications technology (ICT) sector demonstrates the feasibility of the methodology.

The rest of this paper is organized as follows: "Literature review" section reviews related work. "Methodology" section describes the methodology. In "Case study: network dynamics analysis of university-industry collaborations in ICT" section, we present the findings of the empirical study. The last section concludes the study, explains its limitations, and addresses future research directions.

Literature review

Collaborative and knowledge networks in university-industry collaborations research

To understand and promote scientific advances into a productive force, the research on scientific collaboration between universities and firms has long been a focus that has only become more active in the digital era. Sonnenwald (2007) summarizes scientific collaboration as interactions happening within a social context among two or more individuals or higher collectives that facilitates the sharing of the knowledge and accomplishing of tasks concerning a mutually shared superordinate goal. In this sense, collaborative activities are closely related to the interactions and knowledge diffusion of organizations. From a network perspective, Guan and Liu (2016) point out that the scientific collaboration between academia and industry is embedded in both the social networks of organizations and the knowledge networks established by coupled knowledge elements.

Most studies discuss collaborative networks and knowledge networks as two separate strands of inquiry. Collaboration networks are most commonly constructed based on co-authorship/co-patent applications, which gives a bird's eye view of various cooperation at play in a corpus of articles or patents (Liu et al., 2005; Magazinik et al., 2019; Nsanzumuhire & Groot, 2020). For example, Giunta et al. (2016) focus on co-publications in Italy in the field of biopharmaceuticals, outlining several decisive factors that affect university-industry collaborations. Comparatively, only since the last decade, attention has been paid to analyze the content of these interactions from a network perspective. In year 2012, Phelps et al. (2012) conducted a systematic review and analysis of empirical research published on knowledge networks. In the survey, they explained the processes of knowledge creation, diffusion, and absorption, and discussed the use of knowledge networks at length. In recent years, an increasing number of empirical studies have attempted to model knowledge exchanges in collectives as knowledge networks. In research related to university-industry collaborations, these knowledge networks are mainly built from the international patent classification (IPC) codes of co-patent applications of (Guan & Liu, 2016, Chang, 2017), the co-occurrence of publication keywords (Guan et al., 2017), or co-occurring hashtags (Hellsten & Leydesdorff, 2020). This is because these units of analysis offer a very concise representation of the technological features discussed.

Network perspective has and will continue to play an important role in identifying patterns and trends in university-industry collaborations. However, the majority of studies investigate collaborative and knowledge networks in two separate strands, each of which only provide a partial view of the collaborative landscape. Moreover, the knowledge networks are built on simplistic keywords, classification codes, or tags.

What gets overlooked with these tag-based networks is the context of the textual content (Chen et al., 2022). It is difficult to examine actual knowledge structures when only considering straightforward representations based on individual taxonomies. Modelling with the context of collaborations to form richer knowledge networks warrants further research.

Network dynamics studies on collaborative relationships

As a theoretical perspective focusing on the relational structures associated with embedded entities, network approaches have been used to explain a wide range of social phenomena (Borgatti et al., 2009). It is commonly thought that the position of items in a network structure determines an entity's ability to absorb, create, and transfer knowledge (Liefner & Hennemann, 2011). Hence, indicators such as the average shortest distance of the network, the degree distribution, and the aggregation degree have been applied to measure the structural characteristics of collaboration networks (Newman, 2003; Barrat et al., 2004). Most studies apply these indicators to static networks. Yet, in recent years, an increasing number of studies have attempted to describe the dynamic evolution of collaboration networks using streaming-like datasets. For example, Kim et al. (2014) applied node degree, network density, and centrality indicators to describe the structure and evolution characteristics of a cooperation network between the main clusters of the software industry. Zhang et al. (2016) explored the dynamic evolution and characteristics of inter-organizational university-industry collaboration networks using social network analysis from an ego-network perspective. Fischer et al. (2019) selected the 12 most eminent universities in Brazil for the years 1994, 2004, and 2014 and explored the evolution of patenting activity and linkages to industry via a co-assignee network. That said, longitudinal social network studies tend to be costly and time-consuming since multiple static networks need to be constructed and measured (Stadtfeld et al., 2020). Although the trend of the above-mentioned indicators can be modeled and observed to track network dynamics, it is very hard to discover the factors that drive collaborations.

From static to dynamic, research on collaborative network evolution was then extended further, drawing on the statistical inference literature. Research on network dynamics sees social networks as complex adaptive systems and suggests that the naturally accompanying forces operating at the microlevel drive the evolution of a network's macrostructural characteristics (Walter et al., 2005; Snijders et al., 2010). Taking the perspective of network dynamics, Zhang and Luo (2020) turned to stochastic actor-oriented models (SAOMs) to analyze the relationships between network capital, exploitative innovation, and exploratory innovation. Cao et al. (2017) also used SAOMs to examine how the macro structure of a project-based collaborative network evolved over time and how related micro-mechanisms collectively underpin the network's evolution. Hossain et al. (2015) applied exponential random graph models (ERGMs) to explore the micro-level network structures of emergency management networks and their impact on performance. Approaches such as SAOMs and ERGMs for the statistical analysis of social network generation are widely used in the literature. Although statistical analysis of social networks has made considerable advances over the last decade, effective methods of constructing networks using empirical observations and ways to analyze the diverse mechanisms that underlie the forming of these networks has been and will continue to be main focuses for this area.

Embedding approaches in scientific knowledge representation and extraction

Embedding approaches based on neural networks have been in the limelight of scientific text mining research in recent years (Zhai et al., 2021). These approaches are capable of translating high-dimensional sparse vectors into a low-dimensional dense space, providing efficient tools for mapping words or sentences into fixed-length numeric vectors and, in so doing, capturing latent semantics from massive tracts of textual data (Zhang et al., 2018). As one of the most representative and well-recognized embedding techniques using neural networks, the word2vec method has manifested its superiority in analyzing words semantically (Mikolov et al., 2013). Building on the main idea of word2vec, the doc2vec method considers order of words within sentences, going beyond the word level to make it possible to acquire a distributed vector representation at the document level (Le & Mikolov, 2014).

So far, word2vec and doc2vec have achieved fruitful outcomes in scientometric studies. For example, Lenz and Winker (2020) applied doc2vec to technology news articles, extracting topics by clustering document vectors and exploring their diffusion process. Feng (2020) calculated the cosine distance of vectors derived from patent abstracts and proposed an index to quantify the proximity of knowledge. Thijs (2020) regarded scientific papers as a combination of small units, aiming to measure similarity of paragraphs within and between documents. Wittfoth (2019) found a way to identify potential Standard Essential Patents (SEPs) based on the semantic relations between patents. Kim and his colleagues (2020) used doc2vec to generate vectors from both the profiles of startups and the abstracts of patents. They then constructed a framework for selecting promising startups by measuring the technological distance between the two kinds of vectors. Jin et al. (2022) investigated the associations between funding agencies and the topics they fund with a word embedding-enhanced organization-topic network to reveal funding patterns at both the organizational level and the topic level.

To sum up, both the word2vec and doc2vec methods have demonstrated their efficiency in turning massive textual data into dense vectors, which can be further used for clustering, classification, network construction, and so on. Owing to the advantage of retaining the semantic information from a set of documents, these tools will continue to provide new inspiration for scientific text mining research.

Methodology

Framework of collaboration-knowledge dual-layer networks

The challenge of modeling both actor collaborations and the knowledge they produce with a longitudinal dataset requires seeing the collaborations and the knowledge interactions as a multilayer network. In our case, this network is a dual-layer network comprising an organization layer $G_C(V_{Organization}, E_{Organization})$ and a knowledge layer $G_K(V_{Topic}, E_{Topic})$, as shown in Fig. 1. The data used to construct the network is patent data drawn from Derwent Innovation Index database (DII), one of the most important indicators of innovation output. Joint patent applications of academic and industrial organizations represent key evidences of a cooperation between two or more entities (Teng et al., 2021). Hence, two nodes in the organization layer are linked if they have jointly applied for a patent. The organization set is defined as $V_{Organization} = (o_1, o_2, o_3, \dots, o_p)$, where p shows the number

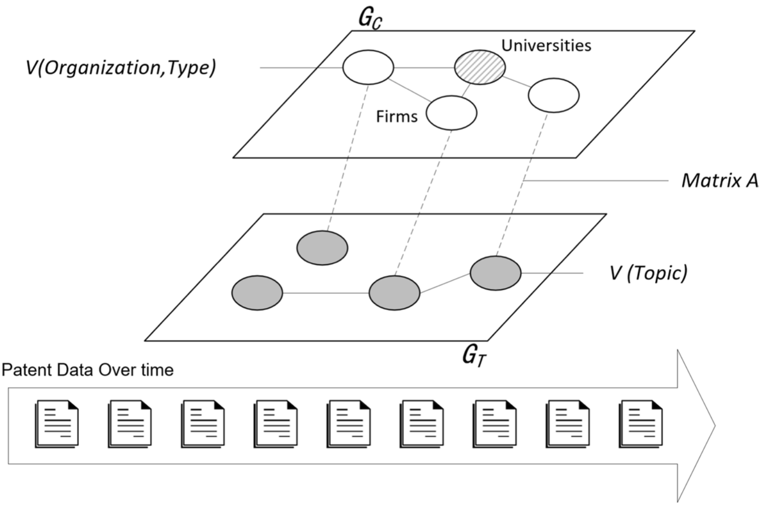

Fig. 1 Framework of the university-industry collaborations dual-layer network modeling collaborative 'actors' interactions and semantic correlations comprehensively

of organizations. To construct the knowledge layer, topics are extracted from the patent titles and abstracts as nodes, and their semantic correlations form the ties. The topic set is defined as $V_{Topic} = (t_1, t_2, t_3, \ldots, t_k)$, where k shows the number of topics. The connections of the two layers can be represented as a $p \times k$ matrix A, reflecting the main topics that each organization is focused on. Modeling dual-layer networks in multiple data collection waves, then allows us to investigate the dynamics that underpin the landscape and the factors contributing to the network's evolution.

Collaboration-knowledge dual-layer network construction with doc2vec

Data pre-processing

As this study applies patents data from DII database to construct collaboration-knowledge dual-layer network, organizations are first taken from the assignee field of the patent data. Universities and firms then need to be identified and tagged based on the following assumptions.

Assumption 1: The university category comprises universities, other higher education providers, research institutes and academic research laboratories. The firm category of organizations consists of industrial enterprises built on the theory of a firm, public enterprise, or non-profit organization that provide products or services to society.

Assumption 2: An assignee is tagged as a university if its name contains the words, "University", "Institution", "School", "College", "Faculty" and so forth. Organizations are tagged as a firm if their names contain "Ltd" (Limited), "Co." (Company), and so on.

To construct the knowledge layer, instead of using concise but oversimplified IPC codes, we extract topics from the titles and abstracts of the patent documents as knowledge elements. To fully leverage the power of topic extraction in exploring semantic structures, the words in the corpora should be lemmatized, then the textual data needs to be cleaned and consolidated before topic extraction. The lemmatization rules we applied included: (1) returning plural nouns to single form; (2) turning inflected forms of verbs back to their stem; (3) returning comparative adjectives to their basic form. In addition, all punctuation, non-alphabetic characters, stop words and common words used in patents should be eliminated (Chen et al., 2021).

Topic extraction

Over the past decade, methods of topic extraction for unstructured text data have been enhanced via the power of modern computing techniques and machine learning algorithms (De Battisti et al., 2015). Our methodology applies one of the most accepted topic modeling techniques to extract topics from reprocessed corpus dataset D–latent Dirichlet allocation (LDA). Each document is considered to be a mixture of topics with a latent topic distribution of ϑ, while each topic is considered to be a mixture of words with a proportion of φ. The n^{th} word in document d is denoted as $w_{d,n}$, the topic assignment for d is denoted as $\vec{Z}_d$, and the corresponding topic distribution is denoted as $\vec{\theta}_d$. LDA assumes that there are K topics that can be denoted as $\vec{\varphi}_{1:k}$, and, as mentioned, each $\vec{\varphi}_k$ is a distribution of words (Heinrich, 2005). The generation process of topics is represented by a joint distribution of random variables (Blei, 2012):

$$p\left(\vec{w}_d, \vec{z}_d, \vec{\vartheta}_d, \phi|\vec{\alpha}, \vec{\beta}\right) = \prod_{n=1}^{N_d} p\left(w_{d,n}|\vec{\varphi}_{z_{d,n}}\right) p\left(z_{d,n}|\vec{\vartheta}_d\right) p\left(\vec{\vartheta}_d|\vec{\alpha}\right) p(\phi|\vec{\beta}), \tag{1}$$

in which α and β are hyper-parameters that control the amount of smoothing applied to the topic and word distributions.

When applying LDA to scientific text mining, assumptions about the number of topics K is set to have attracted much attention. Manually setting K is a convenient option, but this will reduce the reliability of the results. Alternatively, statistical evaluations tend to create a larger number of topics, which can be challenging when it comes to interpreting the results (De Battisti et al., 2015). To balance both concerns, one must make a trade-off between a statistical evaluation and the number of topics that is possible for a human to parse. This can be done by measuring how well a trained model fits the dataset with a perplexity score (Huang et al., 2018), as shown in Eq. (2), while, at the same time, applying an exponential function to describe the interpretation complexity of the model (Chen et al., 2022), as shown in Eq. (3).

$$Perplexity(D) = exp - \frac{\sum_{d=1}^{M} \log\left(p(w)\right)}{\sum_{d=1}^{M} N_d} \tag{2}$$

$$complexity(K) = exp\left(\frac{K - min(K)}{\max(K) - min(K)}\right) \tag{3}$$

Here, N_d in (2) stands for the document length of d in D, which contains a total of M documents, and where $\sum \log(p(w))$ represents the likelihood of the corpus given the trained model with a specific value for K. We prefer a K that can produce comparatively lower perplexity score and a smaller complexity score, which together suggests a lower chance for misrepresentation and better human interpretation. As shown in Eq. (4), the right choice for parameter K is when the sum value of the normalized perplexity and complexity score is at its minimum.

$$\arg\min_{K} f(K) = \frac{perplexity(K) - \min perplexity(K)}{\max perplexity(K) - \min perplexity(K)} + \frac{complexity(K) - \min complexity(K)}{\max complexity(K) - \min complexity(K)} \tag{4}$$

Topic vectorization and network construction

Although LDA can provide a meaningful decomposition of a target corpus, methods based on topic modeling have inherent limitations in detecting similarity ratios via direct distance measurements (Jung & Yoon, 2020). This is especially true with longitudinal data sets. Document embedding techniques can retain the contextual semantics of the textual data. We thus continue vectorize extracted topics via document embedding–more specifically, by using doc2vec, to overcome above mentioned limitation of topic modeling. Doc2vec is an extension of word2vec (Mikolov et al., 2013). It finds fixed-length numeric vector representations of a document by training a document embedding to predict the words in the document (Le & Mikolov, 2014). The goal of word2vec is to maximize the predicted log probability given a sequence of training words $w_{i-k}, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_{i+k}$, in which w_i stands for a target word and k is the window size according to the notations given by Kim et al. (2019) and Zhang et al. (2018). The prediction is generally performed via a multiclass classification formulated with a softmax function as per Eq. (5) below:

$$P\left(w_i | w_{i-k}, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_{i+k}\right) = \frac{e^{y_{w_i}}}{\sum_j e^{y_j}} \tag{5}$$

where each y_j is the j^{th} output value of a feed-forward neural network computed from Eq. (6):

$$y = b + Uh\left(w_{i-k}, \ldots, w_{i+k}; W\right) \tag{6}$$

Here, b, U denotes the bias terms and weight matrix between the hidden layer and the output layer; h and W respectively denote the average of the context words and the word embedding matrix. Following the same main idea of word2vec, in doc2vec, each document is mapped to a unique vector that is represented by a column in a document matrix Doc, while each word is mapped to a unique vector that is represented by a column in the word matrix W. The only difference for doc2vec when formulating the network is adding Doc to the equation below (Kim et al., 2019):

$$y = b + Uh\left(w_{i-k}, \ldots, w_{i+k}; W, Doc\right) \tag{7}$$

For doc2vec, there are two models that can be used to achieve the goal of turning documents into vectors. One is the distributed bag of words (PV-DBOW) model and the other is distributed memory (PV-DM). Although appropriate model selection for doc2vec has long been discussed (Levy et al., 2015; Shahmirzadi et al., 2019), no agreement has been reached on the performance difference between PV-DBOW and PV-DM. Our framework prescribes PV-DM, which Le and Mikolov (2014) report as being superior for training document vectors using the Genism toolkit. Most scientometric studies set the PV-DM hyperparameters based on experience, testing, or reported best practices in the literature (Thijs, 2020; Kim et al., 2017). However, there is no objective consensus on hyperparameter optimization either.

Mendsaikhan et al. (2019) observe that the settings of window size and the subsampling rate only have a slight effect on document similarity tests. But the number of training epochs do make a difference to doc2vec's effectiveness. Curiskis et al. (2020) point out that more epochs are needed for smaller document sizes. Based on existing discussions on these settings, we set the dimension of the document vector γ to 200, the number of epochs to 40, and the window size to 5.

The topic modeling results and the θ matrix are then integrated with the document vectors provided by doc2vec, and K extracted topics are mapped into vectors. As shown in Fig. 2., for a period t, K extracted topics are set as nodes. These represent the core knowledge elements. Their semantic correlations are set as links and, with this, the knowledge network is constructed.

When constructing longitudinal networks, Pearson's correlation coefficient can be used to calculate the similarity between topics across different periods. As shown in Eq. (8), for two connected periods, we have a topic-correlated matrix $TC(T_{p(t)}, T_{q(t+1)})$, in which $T_{p(t)}$ is the topic vector of topic p in the period t and $T_{q(t+1)}$ is the vector of topic q in the period $t+1$. To highlight strong links and to identify the same topics in longitudinal datasets, only correlated values greater than the upper quartile should be kept.

$$TC\left(T_{p(t)}, T_{q(t+1)}\right) = \frac{\sum_{i=1}^{\gamma}\left(T_{p(t),i} - \overline{T_p}\right)\left(T_{q(t+1),i} - \overline{T_{q(t+1)}}\right)}{\sqrt{\sum_{i=1}^{\gamma}\left(T_{p(t),i} - \overline{T_{p(t)}}\right)^2}\sqrt{\sum_{i=1}^{\gamma}\left(T_{q(t+1),i} - \overline{T_{q(t+1)}}\right)^2}} \tag{8}$$

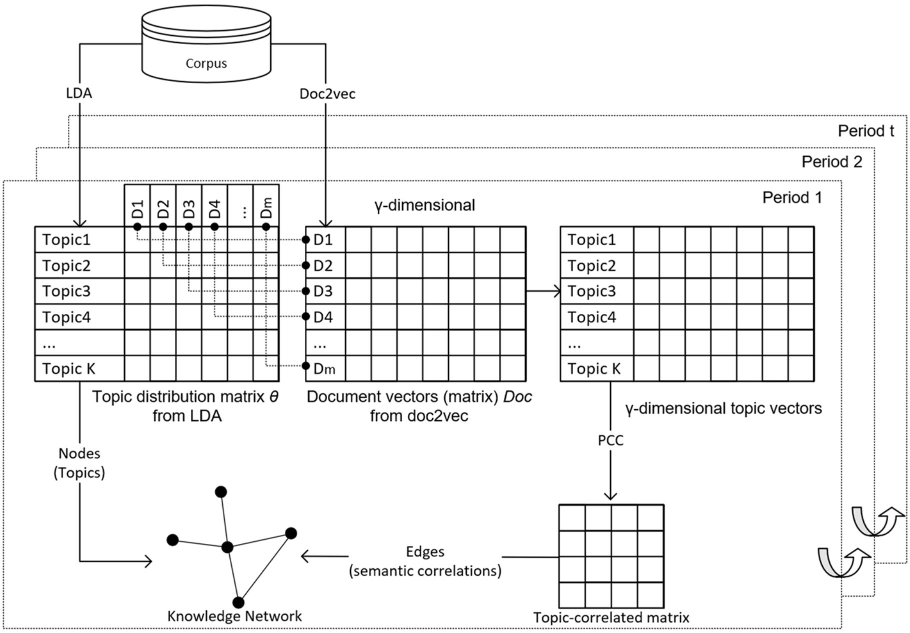

Fig. 2 The process of longitudinal knowledge network construction with topics and topic vectors

Modeling network dynamics

As mentioned, the goal of network construction is to build a collaboration-knowledge dual-layer network consisting of: an organization layer $G_C(V_{Organization}, E_{Organization})$, with tagged organizations as nodes $V_{Organization}$ and patent co-applications as edges $E_{Organization}$; and a knowledge layer $G_K(V_{Topic}, E_{Topic})$ where the extracted topics are the nodes V_{Topic} and their semantic correlations are the edges E_{Topic}. With this network built, one can comprehensively measure both actor collaborations and the knowledge produced, putting a shape to university-industry collaboration dynamics. Constructing a longitudinal analysis, a stochastic actor-oriented model (SAOM) models the network's evolution over different evaluation periods.

Network characteristics analysis

According to Teng et al. (2021), network structures, individual characteristics, and proximity have a significant impact on innovation networking. Network structures reflect the relationships between the elements in a network. Thus, structural characteristics have an important impact on the future interactions of network elements (Yuan & Xu, 2017). For example, Ter Wal and Boschma (2011) find that actors located at the core of a network are better able to enhance their innovation abilities. The characteristics of innovative actors differ in terms of nature, scale, type, and so forth, which affords different roles to the actors in a network. In this vein, Cao et al. (2017) examined network evolution from both the endogenous perspective of network structures and the exogenous perspective of actors' individual characteristics. There has also been some discussion on multi-dimensional proximity in network dynamics (Fitjar and Rodríguez-Pose, 2013; Molina-Morales et al., 2014). Since our case considers knowledge elements and relations formed by the content of interactions between organizations, our focus is on investigating cognitive proximity, following O' Connor et al. (2021).

Several metrics associated with network structures are pertinent to the methodology. The macro structural characteristics include network scale, network density, average path length, and clustering coefficient (Fleming & Frenken, 2007; Gilsing et al., 2008).

Network scale is measured as the total number of nodes (organizations and topics) plus the number of network edges.

Network density describes the connectivity of a network and is computed by dividing the total number of connections by the total number of possible connections with the same number of nodes via $D = E/E_{Max}$ (Kong et al., 2019), where E denotes the actual connections, and E_{Max} stands for all the potential connections. $E_{Max} = n(n-1)/2$, n is the total number of vertices in the target network. A lower network density indicates looser connections between the elements, which implies the behaviors of network elements will receive less influence from the network structure.

Average path length distinguishes negotiable networks from comparatively inefficient ones. The average path length L is calculated as the average length of the shortest path between any two nodes, which can be computed via Eq. (9), in which the distance d_{ij} between node i and node j is the total number of links connecting the shortest path of the two nodes, and where N represents the total number of nodes in the network.

$$L = \frac{2}{N(N-1)} \sum_{i \neq j} d_{ij} \tag{9}$$

The global clustering coefficient measures the clusters across the whole network. A cluster exists if there is a connection between two actors in the network along with a triplet

(Kong et al., 2019). This coefficient is calculated via Eq. (10), where C_i stands for local clustering coefficient, and $C_i = 2E_i/k_i(k_i - 1)$. Here, E_i represents the number of edges for vertex i, and k_i indicates the degree of vertex i (Newman, 2003).

$$C = \frac{\sum_i C_i}{n} \tag{10}$$

Individual characteristics also have important impact on network dynamics from the dual perspective of collaboration and knowledge interactions. In the organization layer, the entities differ in terms of their nature, which is the most basic characteristic of collaborative interactions. In the knowledge layer, the indictors of knowledge diversity and uniqueness are used to measure the topics' characteristics.

Knowledge diversity refers to the number of knowledge elements (topics) associated with the target organization. This is used to evaluate the levels of an organization's knowledge stocks. Organizations with higher knowledge diversity have stronger ability to maintain an R&D advantage (Zhao et al., 2019) According to (Brennecke & Rank, 2017), knowledge diversity can be measured using the number of topics associated with the target organization, as shown in Eq. (11), where if organization i has knowledge of topic j, then $x_{ij} = 1$, and $x_{ij} = 0$ otherwise.

$$Diversity_i = \sum_j x_{ij} \tag{11}$$

Knowledge uniqueness describes the exclusivity of the knowledge stored in an organization, indicating the extent to which this organization has unique knowledge in a specialized field. Based on Brennecke and Rank (2017)'s research, knowledge uniqueness is defined as Eq. (12), in which i indicates the i^{th} organization, j indicates the jth topic, if $x_i = 1$ then organization i has knowledge of topic j, and $x_i = 0$ otherwise.

$$Uniqueness_i = \frac{1}{\sum_j^n \sum_i x_i/n} \tag{12}$$

Cognitive proximity evaluates the extent to which the knowledge of two innovators is cognitively proximate or distant to each other (O' Connor et al., 2021). It encapsulates the viewpoint that firms and universities will be better placed to use externally-sourced knowledge if their knowledge bases are similar (Criscuolo et al., 2018). We use *knowledge proximity* to characterize the cognitive proximity of university-industry collaborations. Knowledge proximity can be calculated via the similarity between the knowledge elements of two organizations in the networks. Given organization i and j, $X_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$ and $X_j = (x_{j1}, x_{j2}, \ldots, x_{jk})$ indicate the topic bases of i and j, in which x_{ik} and x_{jk} represent the frequency of i patents on topic k in the document set. The cosine similarity of X_i and X_j then reveals the knowledge proximity of the two organizations.

Stochastic actor-oriented model

Among the different models for statistically analyzing a network's generation and evolution, stochastic actor-oriented models (SAOMs), exponential random graph models (ERGMs), multiple regression quadratic assignment programs (MRQAPs) and gravity models (GMs) are the most commonly used in the research on innovation network dynamics, network relationship estimation, and economics studies (Block et al., 2019; Teng et al., 2021; Uwe & Bastian, 2014). ERGMs are defined from a graph perspective, and these models dominate the literature on cross-sectional networks. SAOMs are defined from the perspective of transition, and these are mainly used with longitudinal datasets (Block et al., 2019). MRQAP models require a comparatively large number of random permutations for estimation; hence, they are not suited to real-world problems (Uwe & Bastian, 2014), while GM is a classic economics model, but when estimating network data, excessive zero values in the sample will cause deviations in the estimation result (Teng et al., 2021). Facing a longitudinal network analysis task in this research, we thus apply one of the most promising empirical tools for network dynamics analysis, SAOM, to analyze the constructed networks (Snijders et al., 2010). The SAOM model focuses on the mechanisms the underlie network formation while at the same time testing the factors that drive network dynamics (Snijders, 2017).

SAOM uses Markov chain Monte Carlo (MCMC) simulations to estimate the model's parameters, including the rate λ and coefficient β of each influencing factor (t-values are used to test the parameters). (Zhang & Luo, 2020).The first step is to simulate the change rate of the relationships between the organizations. This mainly answers the questions of who will change and what is the corresponding change rate. The second step is to then model how the elements in the network have changed in their states. SAOM relies on a rate function to measure changes in the cooperative relationships for the first step. For the second step, a utility function is set as the organization's objective function, and this objective function is then maximized based on the influence of the network characteristics (Snijders et al., 2010).

The SAOM model assumes that dynamics of the network are affected by the indicators of network structure, individual characteristics, and proximity (Snijders, 2017). Following existing research, network density and transitive triads are the network structure variables; the individual characteristics are knowledge diversity, knowledge uniqueness, and organization type (U or I). Knowledge proximity is measured as the similarity between the topic bases of two organizations (Zhang & Luo, 2020, Giuliani, 2013, Teng et al., 2021). All the variables, their types, and explanations are summarized in Table 1.

Case study: network dynamics analysis of university-industry collaborations in ICT

Data

To showcase how the framework works in action, we conducted a case study in the field of information and communications technology (ICT). ICT is a broad field covering all manner of information and communications technologies, and its gamut continues to evolve. Most importantly, the field has wide-ranging innovation and socio-economic impacts across various parts of the economy; hence, it has attracted much research attention from both the academic and industrial communities (Cecere et al., 2014; Røpke, 2012). Following the definition of ICT in the OECD compendium of patent statistics and the search strategy used by Gao et al. (2019), we retrieved 12,189 university-industry collaborative patents in ICT with China as the assignee country from DII database. The period covered was year 2006–2017.

University-industry collaborative patents were defined as patents with both a university and a firm as the assignees. The timespan applied was the Basic Patent Year, which is the

Table 1 The variables of the SAOM for university-industry collaborations dual-layer network dynamics

Effect	Variables	Definition
Network structure	Network density	The ratio of the total number of connections present and the total number of possible connections
	Transitive triads	The number of ternary closures of innovation actors
Individual characteristics	Organization type	University-industry
	Knowledge diversity	The number of topics associated with the target organization
	Knowledge uniqueness	The strength of the exclusivity of the organization's knowledge stores
Proximity	Knowledge Proximity	The similarity of between two organizations' topic bases

first year that an invention was collected in the DII database. We then pre-processed the dataset following the assumptions in "Methodology" section, tagging all participants as either universities or industrial organizations. Figure 3. presents the annual patent counts from 2006 to 2017 and also shows the corresponding number of universities and firms engaged in these collaborations. From the figure, we can see that ICT is a fast-growing area of innovation with an increase in the number of patents year-on-year. Additionally, the number of organizations involved has grown steadily, with academic collaborators increasing faster than industrial collaborators. The sharp rise in the number of participating universities in 2013 fits the general trend in patent growth.

Constructing the dual-layer networks

To chart the evolution of the collaborations, we constructed three dual-layer networks, one for each of the periods: 2006–2009, 2010–2013, and 2014–2017. Four-year window are consistent with other longitudinal research (Zhang & Luo, 2020). We cleaned and consolidated the corpus then followed the steps for constructing the organization layer of the networks outlined in "Methodology" section Sect. 3. Figure 4. presents the organization layers for the above three periods. The green nodes represent universities and the red nodes represent firms.

Subfigure (a) of Fig. 4. shows that, in the first period, there were only limited number of collaborative clusters. Further, the majority of the core organizations were firms not universities. Firms that stand out include Huawei Technologies Co Ltd (HUAW-C), which essentially has collaborations with all the important universities in the network, e.g., Peking University (UYPK-C), the University of Shanghai Jiaotong (USJT-C), and Tsinghua University (UYQI-C). In the second period, the number of universities and industrial organizations starts to equalize as shown in Subfigure (b). Here, the State Grid Corporation of China (SGCC-C) becomes the most significant node in the network. From 2014 to 2017, both the participants and their connections see a marked rise. As a leading firm, the State Grid Corporation of China (SGCC-C) still holds an important place in Subfigure (c).

To construct the topic layer of these dual-layer networks, we ran multiple LDA experiments with the total number of topics K set to different values for all three periods. Balancing perplexity values and interpretation complexity, the parameter K for periods of 2006–2009, 2010–2013 and 2014–2017, we ultimately set K to 20, 25, and 25. To provide a fine-grained decomposition of the document collection, we followed the common setting in existing research, and set the hyper-parameters to α = 0.5 and β =0.01 (Griffiths & Steyvers, 2004). Then via 10,000 iterations of Gibbs sampling, the model inferred the latent variables and distributions and extracted the topics for each period (Heinrich, 2005). The document vectors were then trained via the Genism toolkit. Based on the review of the hyperparameter settings in "Topic vectorization and network construction" section, we set the dimensions of the document vector, parameter γ, to 200, the number of epochs to 40, and the window size to 5. For each period, we received topic vectors with 200 dimensions (latent features). Next, we constructed the topic-correlated matrixes for each period based on Pearson's correlation coefficients and computed the topic similarity across the different periods. In total, 44 topics were extracted from the longitudinal networks, as listed in Appendix 1.

Figure 5. shows the three topic layers of the networks. Across the full period, the network size has increased, and the links between the topics have become slightly denser, which indicates a strong tendency toward knowledge fusion. Some topics have grown in importance over time–for example, *video processing*. In the first period, this topic only had simple connections at the edge of the network, as shown in Subfigure (a). Yet, in the last period, shown in Subfigure (c), this topic sits at the center of a cluster of topics that includes *wireless transmission*, *storage device*, *data transmission*, and *electric control*. Other topics have vanished through the network's evolution, such as *message path* in 2006–2009 and *generator* 2010–2013. New topics have also emerged like *photovoltaic cable* and *management and scheduling* in the very recent period of 2014–2017.

Modeling the network dynamics

We then model the network dynamics and explore the mechanisms that underpin each of the network's structural characteristics along with the evolution across all three. Table 2 shows the basic statistics of the indicators in the three observation periods. As outlined, the scale of the collaboration network has expanded but the density has decreased. This means newly joining actors are tending to cooperate with only one other organization. In other words, they are failing to establish cooperative relationships with multiple nodes. As

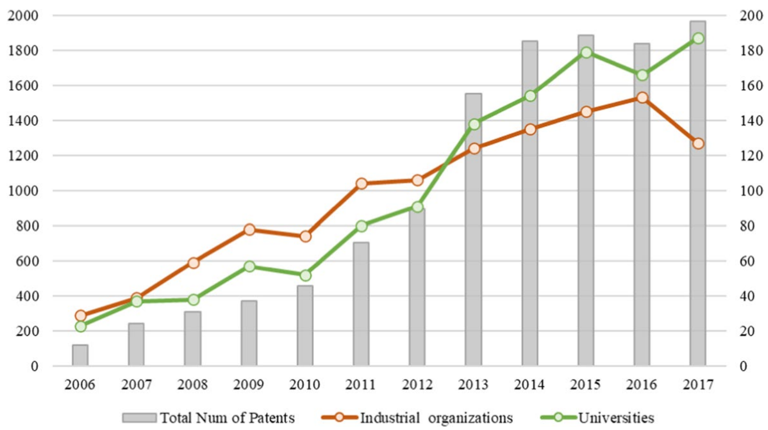

Fig. 3 Number of university-industry collaborative ICT patents with China as the assignee country (2006–2017), plus the number of university and firm assignees

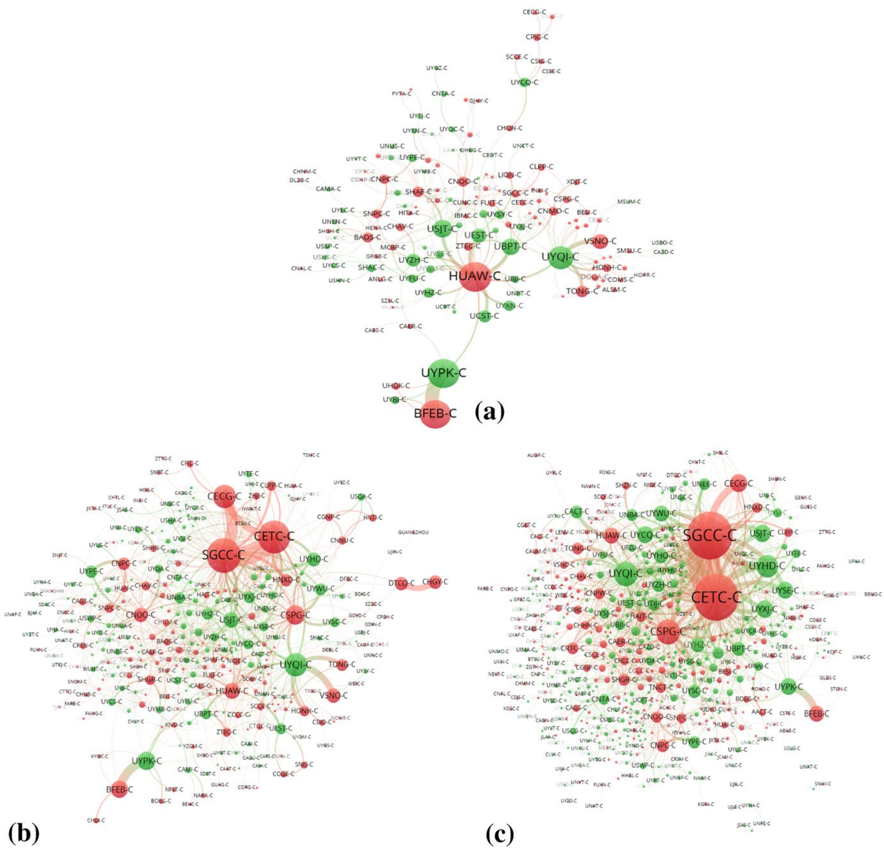

Fig. 4 Collaboration layers of the dual-layer networks for the periods of: **a** 2006–2009; **b** 2010–2013; and **c** 2014–2017

a result, the number of potential edges is much greater than the number of actual connections. In addition, more universities have recently joined these networks than firms, with the collaboration layer showing a 'small world' structure with comparatively a high clustering coefficient and a short average path length.

Turning to the knowledge layer, the network density increased slightly and the network structures show a tendency to evolve around certain core topics during the three periods. The average path length of has increased over time, which means the topic base of the whole field has increased. The clustering coefficient has also increased, indicating an increase in the number of tight topic clusters.

Next, we estimated the SAOM parameters via an R package named RSiena that conducts approximations using MCMC simulations (Ripley et al., 2017). Table 3 reports the parameter estimation results. The t values of all the parameters are significant at $p < 0.001$. At the same time, the $t - ratios < 0.05$ for all parameters, and the overall maximum convergence ratio is 0.0543, which is less than 0.25, signifying good convergence in the model (Ripley et al., 2017).

The rate constants for periods 1 & 2 suggest the speed of organizations creating ties has slowed down. But the estimated positive values of 10.842 and 7.187 indicate there is still a

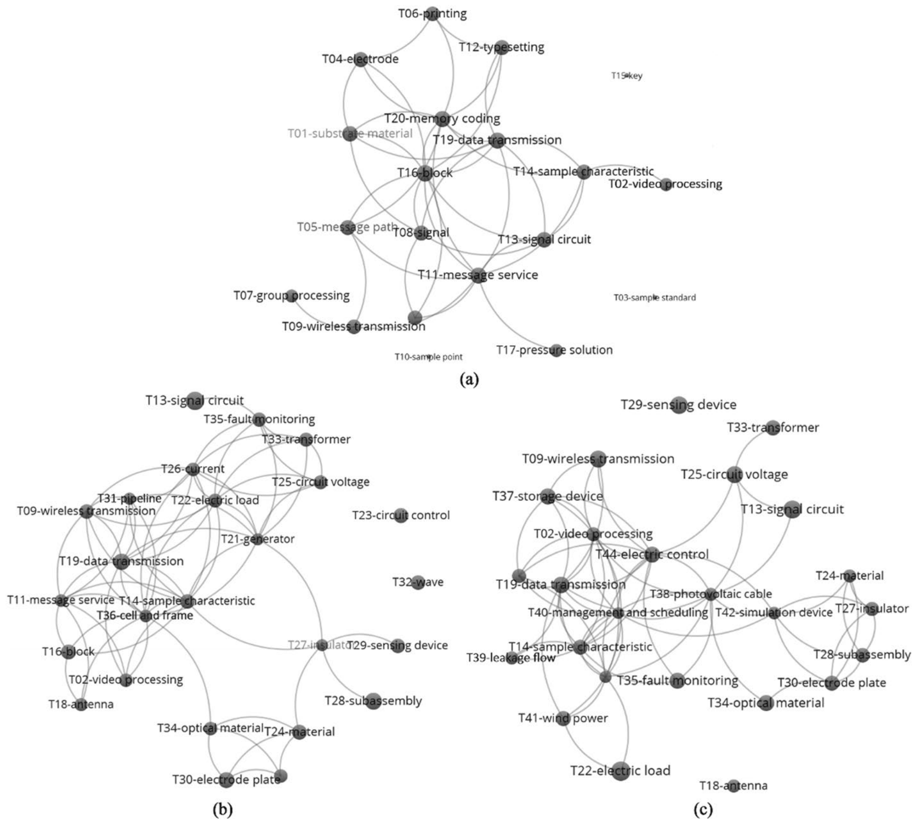

Fig. 5 Knowledge layers of the dual-layer networks for the periods: **a** 2006–2009; **b** 2010–2013; and **c** 2014–2017

significant opportunity for each organization to form new ties in the two periods (Giuliani, 2013). Especially in the period 2006–2009, these organizations had many opportunities to establish collaborations and were more willing to seek partners. This also implies that the collaboration network at this stage was still relatively immature. The negative coefficient of degree i.e., the density, shows no significant influence over the formation of the collaboration network. This suggests that the firms and universities already in collaborations have tended not to establish new ties with other collaborators. Nor have they terminated existing ties with old "friends" (Kalish, 2020). The transitive triads estimation is positive, which reflects that organizations tend to be more willingly to build a tie with another organization that has a mutual collaborator (Zhang & Luo, 2020). From the negative estimation of the organization type variable (0 for firms, 1 for universities), we can tell that firms played a more important role in driving the network's evolution than did universities. In addition, the knowledge diversity and uniqueness variables all have positive values in the parameter estimation. The higher diversity or uniqueness of knowledge that an organization has, the easier it is for that organization to partner with others. The value of knowledge uniqueness ($\beta = 48.049$) also shows that possessing unique a topic base works better than having diverse topics when creating collaborative relationships. The knowledge similarity variable also has a positive estimation result, which shows organizations tend to cooperate with other organizations that have knowledge in similar research topics.

Table 2 Structural statistics in the network's evolution

Collaboration layer	2006–2009	2010–2013	2014–2017
Nodes	203	382	570
Firm nodes	125	198	266
University nodes	78	184	304
Edges	439	1092	1973
Density	0.021	0.015	0.012
Average path length	4.07	3.701	3.394
Clustering coefficient	0.557	0.537	0.505
Knowledge layer	2006–2009	2010–2013	2014–2017
Nodes	20	25	25
Edges	40	66	66
Density	0.211	0.220	0.220
Average path length	2.258	2.616	2.645
Clustering coefficient	0.574	0.659	0.711

Discussion

This empirical study illustrates that the university-industry collaboration activities in China's ICT sector are characterized by sharp growth and a high level of pervasiveness. A considerable number of organizations entered this market over the period, especially from the academic quarter. The dual-layer networks we constructed contain heterogeneous information, and evolutions across the collaboration and knowledge layers presents different cluster characteristics and negotiability. Up until the 2014–2017 period, the collaboration layer was still comparatively immature, while the knowledge layers have shown a tendency to grow around certain key topics, such as the topic of *video processing*. Overall, we find that: (1) The structure of university-industry collaborations in China's ICT sector tends to be loose, and that there is room to improve overall cooperation efficiency. (2) The dynamics of these dual-layer networks are affected by a variety of factors. Transitivity, knowledge uniqueness, knowledge diversity, and knowledge proximity all have positive effects on network evolution. (3) Industry has played a more significant role in driving the network's evolution than academia.

Table 3 The parameter estimation of the SAOM for ICT area in China

	Estimate	Standard error	t-Value
Rate constant r rate (period 1)	10.842	1.330	–
Rate constant r rate (period 2)	7.187	0.394	–
Degree (density)	−2.441***	0.028	−88.388
Transitive triads	0.503 ***	0.005	17.651
Organization type	−0.248 ***	0.073	−3.404
Knowledge similarity	0.636***	3.589	5.412
Knowledge diversity	0.049***	0.005	10.301
Knowledge uniqueness	48.049***	3.589	13.388

†$p<0.1$; *$p<0.05$; **$p<0.01$; ***$p<0.001$; all convergence t ratios < 0.05

Overall maximum convergence ratio 0.0543

In the literature, universities are often seen as central agents in a country's innovation system and its pursuit of technology upgrades (Fischer et al., 2019). However, in our analysis, although many more universities joined the sector than new firms, firms still led network development. To enhance the strategic role of universities in ICT research in China, deeper and wider connections between the two communities are required.

Conclusion, limitation and future work

In this paper, we explored a new perspective on modeling university-industry collaborations and the dynamics at play in these networks. By building a series of longitudinal dual-layer networks spanning collaborative organizations and the knowledge elements they share, we examined the mechanisms that underpin this changing landscape to provide insights into the factors contributing to the network's evolution. Rather than relying on simple four-digits IPC codes to represent knowledge elements, we extended the data source to textual data. Thus, the knowledge elements were derived as topics extracted from the titles and abstracts of patents to provide a better understanding of the technological concepts and ideas being shared. From knowledge discovery perspective, we replaced traditional word-based networks with embedding-derived networks to model network dynamics. The embedding-derived networks provide comprehensive insights into the semantic structure of the topics. Thus, they hold great potential for conducting deep analyses of university-industry collaborations. The results of this investigation provide a more comprehensive understanding of key element identification and knowledge transformation.

Another key aspect of this research was to examine evolutionary characteristics and the driving factors of university-industry collaborations using a stochastic actor-oriented model. Our methodology extends single-layer networks into multilayer networks, which are more reflective of real-world collaboration ecosystems. The big data era has seen a significant reshaping of modes of collaboration, knowledge absorption, and resource exchanges. These mechanisms are now characterized by much higher complexity, which means that one-mode networks may fail to detect the richness of various kinds of interactions. In this context, multilayer networks hold great promise for future innovation management.

Although this paper provides heuristic research that summarizes university-industry collaborations into a dual-layer framework, it has several limitations that need be explored in future research. First, we did not dig deeply into the dynamic mechanisms of how interactions in the network drive knowledge creation, knowledge exchange, and knowledge dissemination. Second, our empirical analysis was limited to a dataset of patents, which is an oversimplification of collaborative activities. Third, our empirical analysis was limited to the ICT field. Therefore, the results may not apply to other industry contexts, although the methodology can be re-applied. We will address the above concerns in future research so as to keep improving the methodology in representing and analyzing the complex system of university-industry collaborations.

Appendix 1

See Table 4

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 72004009).

Table 4 Topics extracted from DII patents for ICT area in China

Global ID	Local ID	Topic	Content (top 5 words)
1	P1-1	Substrate material	Substrate, material, circuit, ink, time
2	P1-2	Video processing	Video, key, motor, power, frame
3	P1-3	Sample standard	Sample, standard, invention, key, switching
4	P1-4	Electrode	Electrode, video, circuit, dielectric, time
5	P1-5	Message path	Message, number, path, invention, module
6	P1-6	Printing	Page, material, typesetting, switching, module
7	P1-7	Group processing	Group, program, processing, mobile_terminal, digital
8	P1-8	Signal	signal, part, pulse, printing, point
9	P1-9	Wireless transmission	Wireless, message, terminal, receiving, mobile
10	P1-10	Sample point	Table, point, sample, line, character
11	P1-11	Message service	Signal, message, service, module, table
12	P1-12	Typesetting	Typesetting, service, character, frame, module
13	P1-13	Signal circuit	Frequency, signal, circuit, control, electronic
14	P1-14	Sample characteristic	Module, character, frequency, mapping, signal
15	P1-15	Key	Key, pressure, client, firewall, server
16	P1-16	Block	Word, block, material, part, electrode
17	P1-17	Pressure solution	Solution, transmission, pressure, request, electronic
18	P1-18	Antenna	Antenna, control, frame, signal, power
19	P1-19	Data transmission	Signal, block, module, transmission, character
20	P1-20	Memory coding	Memory, coding, flow, terminal, time
13	P2-1	Signal circuit	Signal, circuit, frequency, processing, sensor
21	P2-2	Generator	Generator, voltage, standard, instrument, transient
22	P2-3	Electric load	load, power, frequency, Electric, electric_power
23	P2-4	Circuit control	control, protection, Direct_current, voltage, memory
24	P2-5	Material	Material, gas, sample, metal, sampling
25	P2-6	Circuit voltage	Line, circuit, voltage, wire, measuring
9	P2-7	Wireless transmission	Wireless, grid, main, collecting, platform

Table 4 (continued)

Global ID	Local ID	Topic	Content (top 5 words)
26	P2-8	Current	current, character, voltage, Prediction, coil
18	P2-9	Antenna	antenna, base_station, Feedback, uplink, carrier
27	P2-10	Insulator	oil, insulator, word, agent, key
28	P2-11	Subassembly	Pressure, part, sensor, rod, valve
11	P2-12	Message service	Terminal, message, server, client, internet
29	P2-13	Sensing device	Temperature, measuring, water, pipe, heat
30	P2-14	Electrode plate	electrode, surface, arranged, plate, substrate
1	P2-15	Substrate material	Substrate, material, silicon, color, ink
31	P2-16	Pipeline	Interface, pipeline, gas, circuit, management
14	P2-17	Sample characteristic	set, characteristic, wind, Generating, scheduling
32	P2-18	Wave	wave, laser, point, Dimensional, seismic
33	P2-19	Transformer	Voltage, transformer, power, current, power_supply
16	P2-20	Block	Block, matrix, vector, relay, number
19	P2-21	Data transmission	service, power, line, processing, database
34	P2-22	Optical material	Organic_electroluminescent, material, compound, light_emitting, electron
2	P2-23	Video processing	Video, ray, beam, scanning, detector
35	P2-24	Fault monitoring	Voltage, fault, point, current, measuring
36	P2-25	Cell and frame	frame, cell, virtual, light, terminal
14	P3-1	Sample characteristic	Pulse, characteristic, wind, step, point
37	P3-2	Storage device	storage, fault, control, camera, vehicle
24	P3-3	Material	Hole, oil, corrosion, steel, material
30	P3-4	Electrode plate	Plate, sensor, voltage, electrode, vibration
38	P3-5	Photovoltaic cable	Photovoltaic, cable, identification, time, insulator
11	P3-6	Message service	terminal, service, topology, type, supply
39	P3-7	Leakage flow	Flow, electric, leakage, quality, sample

Table 4 (continued)

Global ID	Local ID	Topic	Content (top 5 words)
29	P3-8	Sensing device	Temperature, pressure, gas, pipe, sensor
9	P3-9	Wireless transmission	Signal, intelligent, wireless, electric, collecting
27	P3-10	Insulator	Ring, part, insulation, oil, sleeve
40	P3-11	Management and Scheduling	Management, variable, scheduling, type, binary
41	P3-12	Wind power	Wind, electric, group, time, optimization
42	P3-13	Simulation device	Load, shell, switch, branch, rock
43	P3-14	Reliability evaluation	Reliability, probability, step, electromagnetic_transient, requirement
2	P3-15	Video processing	Video, light, table, box, processing
13	P3-16	Signal circuit	signal, circuit, electric, Voltage, sensor
22	P3-17	Electric load	grid, load, electric, operation, risk
44	P3-18	Electric control	Control, electric, operation, fault, bus
25	P3-19	Circuit voltage	Electric, voltage, circuit, current, direct_current
33	P3-20	Transformer	Transformer, voltage, grid, transformer_substation, state
19	P3-21	Data transmission	Cable, transmission, electric, digital, operation
34	P3-22	Optical material	Light, surface, insulator, optical_fiber, part
28	P3-23	Subassembly	metal, platform, frame, rod, type
18	P3-24	Antenna	Antenna, surface, excitation, cluster, short_term
35	P3-25	Fault monitoring	Fault, line, point, frequency, current

References

Akhtar, P., Khan, Z., Rao-Nicholson, R., & Zhang, M. (2019). Building relationship innovation in global collaborative partnerships: Big data analytics and traditional organizational powers. *R&D Management, 49*, 7–20.

Barrat, A., Barthélemy, M. & vespignani, A. 2004. Weighted evolving networks: Coupling topology and weight dynamics. *Physical review letters,* 92**,** 228701.

Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the Acm, 55*, 77–84.

Block, P., Stadtfeld, C., & Snijders, T. A. B. (2019). Forms of dependence: Comparing SAOMs and ERGMs from basic principles. *Sociological Methods & Research, 48*, 202–239.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*, 892–895.

Brennecke, J., & Rank, O. (2017). The firm's knowledge network and the transfer of advice among corporate inventors–a multilevel network study. *Research Policy, 46*, 768–783.

Cao, D., Li, H., Wang, G., Luo, X., Yang, X., & Tan, D. (2017). Dynamics of project-based collaborative networks for BIM implementation: Analysis based on stochastic actor-oriented models. *Journal of Management in Engineering, 33*, 04016055.

Cecere, G., Corrocher, N., Gossart, C., & Ozman, M. (2014). Technological pervasiveness and variety of innovators in Green ICT: A patent-based analysis. *Research Policy, 43*, 1827–1839.

Chang, S.-H. (2017). The technology networks and development trends of university-industry collaborative patents. *Technological Forecasting and Social Change, 118*, 107–113.

Chen, H., Jin, Q., Wang, X., & Xiong, F. (2022). Profiling academic-industrial collaborations in bibliometric-enhanced topic networks: A case study on digitalization research. *Technological Forecasting and Social Change, 175*, 121402.

Chen, H., Wang, X., Pan, S., & Xiong, F. (2021). Identify topic relations in scientific literature using topic modeling. *IEEE Transactions on Engineering Management, 68*, 1232–1244.

Criscuolo, P., Laursen, K., Reichstein, T., & Salter, A. (2018). Winning combinations: Search strategies and innovativeness in the UK. *Industry and Innovation, 25*, 115–143.

Curiskis, S. A., Drake, B., Osborn, T. R., & Kennedy, P. J. (2020). An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. *Information Processing & Management, 57*, 102034.

de Battisti, F., Ferrara, A., & Salini, S. (2015). A decade of research in statistics: A topic model approach. *Scientometrics, 103*, 413–433.

Feng, S. J. (2020). The proximity of ideas: An analysis of patent text using machine learning. *PLoS ONE, 15*, 19.

Fischer, B. B., Schaeffer, P. R., & Vonortas, N. S. (2019). Evolution of university-industry collaboration in Brazil from a technology upgrading perspective. *Technological Forecasting and Social Change, 145*, 330–340.

Fitjar, R. D., & Rodríguez-pose, A. (2013). Firm collaboration and modes of innovation in Norway. *Research Policy, 42*, 128–138.

Fleming, L., & Frenken, K. (2007). The evolution of inventor networks in the silicon valley and Boston regions. *Advances in Complex Systems, 10*, 53–71.

Gao, X., Qi, G., & Cao, J. (2019). Influence of openness on the innovation performance in the industry–university–research cooperation innovation networks. *Science Research Management, 40*, 231–240.

Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., & van den Oord, A. (2008). Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy, 37*, 1717–1731.

Giuliani, E. (2013). Network dynamics in regional clusters: Evidence from Chile. *Research Policy, 42*, 1406–1419.

Giunta, A., Pericoli, F. M., & Pierucci, E. (2016). University-Industry collaboration in the biopharmaceuticals: The Italian case. *Journal of Technology Transfer, 41*, 818–840.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 5228–5235.

Guan, J., & Liu, N. (2016). Exploitative and exploratory innovations in knowledge network and collaboration network: A patent analysis in the technological field of nano-energy. *Research Policy, 45*, 97–112.

Guan, J., Yan, Y., & Zhang, J. J. (2017). The impact of collaboration and knowledge networks on citations. *Journal of Informetrics, 11*, 407–422.

Guan, J., & Zhao, Q. (2013). The impact of university–industry collaboration networks on innovation in nanobiopharmaceuticals. *Technological Forecasting and Social Change, 80*, 1271–1286.

Heinrich, G. 2005. Parameter estimation for text analysis. Technical report

Hellsten, I., & Leydesdorff, L. (2020). Automated analysis of actor–topic networks on twitter: New approaches to the analysis of socio-semantic networks. *Journal of the Association for Information Science and Technology, 71*, 3–15.

Hossain, L., Hamra, J., Wigand, R. T., & Carlsson, S. (2015). Exponential random graph modeling of emergency collaboration networks. *Knowledge-Based Systems, 77*, 68–79.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach. *Management Science, 64*, 2833–2855.

Huang, M. H., Yang, H. W., & Chen, D. Z. (2015). Industry-academia collaboration in fuel cells: A perspective from paper and patent analysis. *Scientometrics, 105*, 1301–1318.

Jin, Q., Chen, H., Wang, X., Ma, T., & Xiong, F. (2022). Exploring funding patterns with word embedding-enhanced organization–topic networks: A case study on big data. *Scientometrics, in Press:* https://doi.org/10.1007/s11192-021-04253-x

Jung, S., & Yoon, W. C. (2020). An alternative topic model based on common interest authors for topic evolution analysis. *Journal of Informetrics, 14*, 101040.

Kalish, Y. (2020). Stochastic Actor-Oriented Models for the Co-Evolution of Networks and Behavior: An Introduction and Tutorial. *Organizational Research Methods, 23*, 511–534.

Kim, D., Seo, D., Cho, S., & Kang, P. (2019). Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. *Information Sciences, 477*, 15–29.

Kim, H. K., Kim, H., & Cho, S. (2017). Bag-of-concepts: Comprehending document representation through clustering words in distributed representation. *Neurocomputing, 266*, 336–352.

Kim, H. J., Kim, T. S., & Sohn, S. Y. (2020). Recommendation of startups as technology cooperation candidates from the perspectives of similarity and potential: A deep learning approach. *Decision Support Systems, 130*, 12.

Kim, H. D., Lee, D. H., Choe, H., & Seo, I. W. (2014). The evolution of cluster network structure and firm growth: A study of industrial software clusters. *Scientometrics, 99*, 77–95.

Kong, X., Shi, Y., Yu, S., Liu, J., & Xia, F. (2019). Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications, 132*, 86–103.

Le, Q. & Mikolov, T. Distributed representations of sentences and documents. Proceedings of the 31st international conference on machine learning (ICML-14), 2014. 1188–1196.

Lenz, D. & Winker, P. 2020. Measuring the diffusion of innovations with paragraph vector topic models. *Plos One,* 15.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics, 3*, 211–225.

Liefner, I., & Hennemann, S. (2011). Structural holes and new dimensions of distance: The spatial configuration of the scientific knowledge network of China's optical technology sector. *Environment and Planning A, 43*, 810–829.

Liu, X., Bollen, J., Nelson, M. L., & van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing & Management, 41*, 1462–1480.

Magazinik, A., Mäkinen, S. J., Lasheras, N. C., Bedolla, J. S. & Saari, U. Research-industry collaboration: A review of the literature on evaluation methods and motivations. 2019 portland international conference on management of engineering and technology (PICMET), 25–29 Aug. 2019 2019. 1–19.

Mao, C. F., Yu, X. Y., Zhou, Q., Harms, R., & Fang, G. (2020). Knowledge growth in university-industry innovation networks–results from a simulation study. *Technological Forecasting and Social Change, 151*, 9.

Mendsaikhan, O., Hasegawa, H., Yamaguchi, Y. & Shimada, H. Identification of cybersecurity specific content using the Doc2Vec language model. 2019 IEEE 43rd annual computer software and applications conference (COMPSAC), 15–19 Jul 2019 2019. 396–401.

MIKOLOV, T., SUTSKEVER, I., CHEN, K., CORRADO, G. S. & DEAN, J. 2013. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems,* 3111–3119.

Molina-Morales, F. X., GARCíA-Villaverde, P. M., & Parra-Requena, G. (2014). Geographical and cognitive proximity effects on innovation performance in SMEs: A way through knowledge acquisition. *International Entrepreneurship and Management Journal, 10*, 231–251.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Review, 45*, 167–256.

Nsanzumuhire, S. U., & Groot, W. (2020). Context perspective on university-industry collaboration processes: A systematic review of literature. *Journal of Cleaner Production, 258*, 120861.

O' Connor, M., Doran, J., & Mccarthy, N. (2021). Cognitive proximity and innovation performance: Are collaborators equal? *European Journal of Innovation Management, 24*, 637–654.

Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks. *Journal of Management, 38*, 1115–1166.
Ripley, R., Snijders, T. A. B., Boda, Z., VöRöS, A. & Preciado, P. 2017. Manual for SIENA version 4.0.: Oxford: University of Oxford, Department of Statistics.
RøPKE, I. (2012). The unsustainable directionality of innovation–The example of the broadband transition. *Research Policy, 41*, 1631–1642.
Roth, C., & Cointet, J.-P. (2010). Social and semantic coevolution in knowledge networks. *Social Networks, 32*, 16–29.
Shahmirzadi, O., Lugowski, A. & Younge, K. Text similarity in vector space models: a comparative study. 2019 18th IEEE international conference on machine learning and applications (ICMLA), 2019. IEEE, 659–666.
Snijders, T. A. B., (2017) Stochastic Actor-Oriented Models for Network Dynamics. *In:* Fienberg, S. E. (ed.) *Annual review of statistics and its application, vol 4.*
Snijders, T. A. B., van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32*, 44–60.
Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology, 41*, 643–681.
Stadtfeld, C., Snijders, T. A. B., Steglich, C., & van Duijn, M. (2020). Statistical power in longitudinal network studies. *Sociological Methods & Research, 49*, 1103–1132.
Teng, T. W., Cao, X. Z. & Chen, H. T. (2021) The dynamics of inter-firm innovation networks: The case of the photovoltaic industry in China. *Energy Strategy Reviews,* 33.
ter Wal, A. L. J., & Boschma, R. (2011). Co-evolution of firms, industries and networks in space. *Regional Studies, 45*, 919–933.
THIJS, B. (2020). Using neural-network based paragraph embeddings for the calculation of within and between document similarities. *Scientometrics, 125*, 835–849.
Uwe, C., & Bastian, R. (2014). International research networks in pharmaceuticals: Structure and dynamics. *Research Policy, 43*, 333–348.
Walter,W., Powell Douglas, R., White, Kenneth W., Koput & Jason Owen-Smith, 2005. Network dynamics and field evolution: The Growth of interorganizational collaboration in the life sciences. American Journal of Sociology, 110, 1132-1205
Wang, C.-H., & Hsu, L.-C. (2014). Building exploration and exploitation in the high-tech industry: The role of relationship learning. *Technological Forecasting and Social Change, 81*, 331–340.
Wittfoth, S. 2019. Identification of probable standard fissential patents (SEPs) based on semantic analysis of patent claims. In: Kocaoglu, D. F., Anderson, T. R., Kozanoglu, D. C., Niwa, K. & Steenhuis, H. J. (eds.) *2019 Portland international conference on management of engineering and technology.* New York: Ieee.
Yan, Y., & Guan, J. (2018). Social capital, exploitative and exploratory innovations: The mediating roles of ego-network dynamics. *Technological Forecasting and Social Change, 126*, 244–258.
Yuan, J. F., & Xu, Z. (2017). Research on the structural characteristics and evolution of industry-university–research cooperation networks in China: Based on analysis of patent data from 1985 to 2013 years. *Chinese Journal of Management, 14*, 1024–1032.
Zhai, Y., Ding, Y., & Zhang, H. (2021). Innovation adoption: Broadcasting versus virality. *Journal of the Association for Information Science and Technology, 72*, 403–416.
ZHANG, Z. G. & LUO, T. Y. 2020. Network capital, exploitative and exploratory innovations-from the perspective of network dynamics. *Technological Forecasting and Social Change,* 152.
Zhang, R. J., & Ye, F. Y. (2020). Measuring similarity for clarifying layer difference in multiplex ad hoc duplex information networks. *Journal of Informetrics, 14*, 100987.
Zhang, Y., Chen, K., Zhu, G., Yam, R. C. M., & Guan, J. (2016). Inter-organizational scientific collaborations and policy effects: An ego-network evolutionary perspective of the Chinese academy of sciences. *Scientometrics, 108*, 1383–1415.
Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H. S., & Zhang, G. Q. (2018). Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. *Journal of Informetrics, 12*, 1099–1117.
Zhao, L. M., Zhang, H. H., & Wu, W. Q. (2019). Cooperative knowledge creation in an uncertain network environment based on a dynamic knowledge supernetwork. *Scientometrics, 119*, 657–685.