Collaboration and Knowledge Networks: a Framework on Analyzing Evolution of University-industry Collaborative Innovation

Hongshu Chen\textsuperscript{1}[0000-0002-0893-1817] and Xinna Song\textsuperscript{1}

\textsuperscript{1} School of Management and Economics, Beijing Institute of Technology, China
Hongshu.Chen@bit.edu.cn

Abstract. Collaborations of universities and firms provide a key pathway for innovation. In recent years, interactions between the two communities have been reshaped with much higher complexity due to the enhanced ability of creating, disseminating and exchanging knowledge in big data era. This short paper aims to improve the framework of modeling interactions in University-Industry collaboration by building both participant cooperation network and knowledge network for re-recognizing the patterns, characteristics and evolution trend of university-industry collaborative innovation. The proposed framework integrates techniques of bibliometrics, complex network analysis, and text mining to reveal the evolution of both participants interactions and their collaborative knowledge structure in a two-layers multiplex network. Network analytic metrics are selected to provide comprehensive insight on structural properties and characteristics. Finally, the industry of information and communications technology (ICT) is selected to provide an empirical case study to examine the feasibility of the framework.

Keywords: Collaborative Network, Knowledge Network, University-industry Collaboration.

1 Introduction

As an important part of innovation-driven development strategy, University-Industry Collaboration (UIC) is one of the most important ways of boosting innovative capability of organizations, countries or higher-level collectives [1]. In recent years, as the UI connections has been reshaped with much higher complexity due to the enhanced ability of creating, disseminating and exchanging knowledge in big data era, an increasing number of studies have examined the interactive mode of UIC from a network perspective [2-4]. Under such circumstances, how to re-recognize the patterns, characteristics and evolution trend in a new system of complex network is of more theoretical and practical significance, for not only revealing the potential cooperative opportunity, but also preparing the collaboration while planning the strategic innovation.
As a matter of fact, UIC innovations are doubly embedded in social networks of participants and knowledge networks constituted by coupling among knowledge elements [5]. The nodes and links in both social networks of organizations and knowledge networks not only reveal the ‘actor’ interactions but also explain the processes and trend of knowledge interaction and absorption. Existing research has provided promising analysis using research article co-authorship, patents co-application, grants co-fund and research contracts cooperation to model the interactions of ‘actors’ in UIC via social network analysis [6]. However, comparatively less attention has been paid in analyzing the content of these interactions, let alone profiling the collaborative patterns in the dual-networks perspective of actors and content.

This short paper aims to improve the framework of modeling UIC interactions by building both collaboration networks and knowledge networks for future research in mechanism and opportunity discovery. The proposed framework integrates techniques of bibliometrics, complex network analysis, and text mining to reveal the evolution of both UI cooperation and their knowledge transformation in a multiplex network. Network metrics are selected to provide comprehensive insight on structural properties and characteristics. Finally, the industry of Information and Communications Technology (ICT) is selected to provide an empirical case study to examine the feasibility of the framework.

The rest of this paper is organized as follows: Section 2 reviews related work in existing research. Section 3 describes the full process and modules of constructing the framework. In Section 4, we present the findings observed in the empirical study using ICT patents via the multiplex network analysis. The last section concludes this study, explains the limitation and addresses future research directions.

2 Literature review

2.1 Collaborative and knowledge networks in UIC

To understand and then promote scientific advances transformation into productive forces, the research on scientific collaboration between universities and firms has long been a joint focus, and become more active in big data era. Sonnenwald [7] summarized 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 practice, both academia and industry can garner substantial benefits from scientific collaborations. For universities, collaboration with firms can provide opportunities to gain cutting edge insight by complementing their theory with practice [8, 9]; for industrial collaborators, it is beneficial to broaden knowledge scope by contacting with universities and enhancing innovative capability [10].

The scientific collaboration between UI are embedded in both social networks of organizations and knowledge networks established by coupling among collaborative content [5]. Existing studies discuss the two networks in two separate strands. For collaboration networks formed by participants, co-authorship/co-application network analysis, have mainly been adopted to get a bird’s eye view of the structure of collab-
oration and the status of authors/assignees for article and patent data [11, 12]. Comparatively, less attention has been paid to analyzing the content of these interactions from a network perspective. In year 2012, Phelps, Heidl and Wadhwa [13] refer to the networks social relationships constitute in explaining the processes of knowledge creation, diffusion, absorption and use as ‘knowledge networks’, and conducted a systematic review and analysis of empirical research published on knowledge networks. In recent years, an increasing number of empirical studies model knowledge exchange in social relationships among collectives as knowledge networks [5, 13]. These knowledge networks are mainly built with co-application of international patent classification codes [4, 5] and co-occurrence publication keywords [14].

Network perspective plays an important role in identifying patterns of collaboration. The nodes and links in both social networks of organizations and knowledge networks can present collaborative ‘actors’ and their interactions also knowledge ‘elements’ and their semantic relations. However, the two type of networks in existing studies are analyzed in two separate strands in majority cases, providing only partial view of the UIC, thus warrant further research.

2.2 Evolution of collaboration in research cooperation

The position of items in a complex network structure determines its ability to absorb, create, and transfer knowledge to a certain extent [15]. Based on complex networks, Newman [16] and Barrat, Barthélémy and Vespignani [17] applied the average shortest distance of the network, degree distribution, aggregation degree and other indicators to measure the structural characteristics of collaboration network, and studied the statistics of complex networks. With the assistance of these indicators, structure properties of networks can be revealed. Furthermore, an increasing number of studies described the temporal and spatial evolution of cooperative networks using streaming-like datasets. Fischer, Schaeffer and Vonortas [2] selected the twelve most eminent universities in Brazil for the years 1994, 2004 and 2014 and explored the evolution of patenting activity and linkages to industry via co-assignee network. Kim, Lee, Choe and Seo [18] applied node degree, network density, and centrality indicators to describe the structure and evolution characteristics of the cooperation network between the main clusters of the software industry.

However, existing UIC research mainly focused on depicting the structural properties of participants’ collaborative networks (i.e., individuals, teams and institutions), or spatial networks (cities and countries). The evolution of knowledge networks in UIC context was seldom studied jointly. The evolution analysis of collaboration networks and knowledge networks was discussed the in two separate strands as well.

3 Methodology

3.1 Collaboration and knowledge networks construction

The challenge of analyzing both collaborative and knowledge interactions in UIC requires combining design and methods from social and knowledge network analysis.
In this research, we apply patents data from frequently used Derwent Innovation Index database (DII) to construct collaboration and knowledge networks. The data-preprocessing assumptions and modules can also be applied to scientific articles data from Web of Science (WoS) databases and so forth. We first pre-process the assignee organization field in patents, with assumptions for Natural Language Processing (NLP) and data categorization purposes. Universities and industrial organizations are recognized and tagged based on these definitions.

Assumption 1: industrial organizations are those were built on the theory of the firm, or public enterprises, or non-profit organizations, which provide products or services to the society. Universities and higher education providers are academic organizations. We also include research institutions and academic research laboratories, which dedicated to education and scientific research to this category.

Assumption 2: We tag assignee organizations in patents with terms indicating academic feature as university-side organizations: for example, “University”, “Institution”, “School”, “College”, “Faculty” and so forth. Meanwhile, all organizations are given an industrial tag if their names containing “Ltd” (Limited), “Co.” (Company), and so on.

This study constructs a multiplex network consisting of two layers: organization layer $G_c(V_{org}, E)$ and knowledge layer $G_k(V_{topic}, E)$. For the organization layer, we set tagged organizations as nodes and drew their collaboration relations as links. For knowledge network, it has four-digits international patent classification codes (IPC) defined by the World Intellectual Property Organization (WIPO) as vertices and their co-occurrence relations as ties to quantitatively measuring semantic structure. The IPC code is one of the most straightforward proxies considering the availability of data and has been set as knowledge ‘elements’ in existing research of network construction [5]. Modeling the multiplex networks in annual patent datasets, we can then track the evolution of UIC in the target area, and further analyze future trends based on the historical evolution of the networks. Fig. 1 shows the framework of UIC multiplex network.

Fig. 1. Framework of UIC multiplex network modeling collaborative ‘actors’ interactions and knowledge ‘elements’ relations comprehensively.
3.2 Network analysis metrics

Network structure reflects the relationship between elements of the network, and the characteristics of the structure have important impact on the future interactions of network elements [19]. In this paper, we mainly discuss the change of network cluster characteristics and negotiability in the process of the multiplex network evolution.

We applied network density and global clustering coefficient to depict the cluster characteristics of both layers of UIC multiplex network. As one of the most important global metrics, network density describes the connectivity of a network, which is calculated by dividing the total number of connections present by the total number of possible connections with the same number of nodes [20], as shown in equation (1).

\[ D = \frac{E}{E_{\text{Max}}} \]  

in which \( E \) denotes the actual connections, and \( E_{\text{Max}} \) stands for all the potential connections. \( E_{\text{Max}} = \frac{n(n - 1)}{2} \), \( n \) is the total number of vertices in the target network. Smaller network density indicates looser connections between the elements, which implies the behaviors of network elements will receive less influence from the network structure.

Global clustering coefficient is calculated based on local clustering coefficient of each node, measuring the clusters of the whole network, which can be computed via equation (2)[16].

\[ C = \frac{\sum C_i}{n} \]

in which \( C_i \) stands for local clustering coefficient, and \( C_i = \frac{2E_i}{k_i(k_i - 1)} \). Here \( E_i \) represents the number of edges for vertex \( i \), and \( k_i \) indicates the degree value of vertex \( i \).

In this study, we applied the metric of average path length to distinguish 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 equation (3).

\[ L = \frac{2}{N(N-1)} \sum d_{ij} \]

in which the distance \( d_{ij} \) between node \( i \) and node \( j \) depicts 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.

4 Empirical case study: evolution analysis of UIC in ICT area

4.1 Data

We select Information and Communications Technology (ICT) field as the target area to conduct an empirical case study in this section. ICT has become an extensional
description that covers information technology and communications, and its concept keeps evolving and broadening in recent years. It can be applied in different domains with wide-ranging innovative and socio-economic impacts across various parts of the economy, thus attracts continuous research attention from both academic and industrial communities [21, 22]. This paper followed the definition on ICT industry in OECD compendium of patent statistics, retrieved 25,749 U-I collaborative ICT patents with China as assignee country from the Derwent Innovation Index database (DII), during the time period of 2006-2017. We pre-processed the dataset followed the assumptions in Section 3 and tagged all universities (academic organizations) and firms.

Based on the first time an invention is collected in DII, Fig. 2 presents the annual number of patents in each year from 2006 to 2017, and also shows the corresponding numbers of universities and firms engaged in UI collaborations every year. The general trend of patent grant of Chinese ICT industry is growing continuously, which confirms fast growing research and development interests of ICT in recent years. Both participant universities and industrial organizations are growing steadily. The growth trend of academic collaborators is more flatten, while the rising tendency of participating firms received marked rise from year 2012 to 2017, which basically fit the general trend of patents growth.

4.2 UIC Multiplex network construction

We constructed two layers of the UIC multiplex network to model the collaboration and knowledge interactions for ICT area. For the organization layer, nodes were tagged organizations, while ties were set as their collaboration relations. For

---

1 We define U-I collaborative patents as patent that have both universities and firms as assignees. The field of timespan setting we applied for patent retrieval is Basic Patent Year, which is the first time an invention is collected in the DII.
knowledge network, vertices were four-digits IPC codes and links were their co-occurrence relations. As one of the most straightforward proxies of technological knowledge considering the availability of data, IPC nodes provided concise description of knowledge elements in ICT. Further semantic analytics can be conducted using patent abstracts or claims with scientific text mining techniques, we will discuss the possible improvement for knowledge element extraction in the discussion, research limitations and future work.

Although patents are not typical dynamic data in the narrower sense. Patents are granted on a continual basis. Two-layers multiplex networks based on annual patent data were constructed to present the historical evolution of UIC in ICT. Fig. 3 presented the collaboration layer of multiplex networks in year 2006 and year 2017, in which all the academic organizations were marked in yellow and all the industrial organization were highlighted in purple.

![Collaboration networks in year 2006 and year 2017: subfigure (a) presents the collaboration network in year 2006; subfigure (b) shows the collaboration network in year 2017.](image)

We can observe from Fig. 3 subfigure (a) that in year 2006, there were only limited number of ICT organizations collaborative clusters. Majority of the core organizations were universities rather than firms, for example, Peking University (UYPK-C), University of Shanghai Jiaotong (USJT-C) and Tsinghua University (UYQI-C). At this stage, the technological collaboration of ICT was mainly leading by academic organization. The only firms stood out in year 2006 was Huawei Technologies Co Ltd (HUAW-C). In year 2017, both the participants and their connections, or we say nodes and ties, of the of UIC have received marked rise. As shown in Fig. 3 subfigure (b), as a leading firm, State Grid Corporation of China (SGCC-C) became the most significant node in the collaborative network. Although universities like Tsinghua University (UYQI-C) still served as the central role on the edge of the network, an increasing number of firms started to take the core position in the clusters.
From year 2006 to 2017, the size of corresponding knowledge network grew substantially. Fig. 4 selected and presented knowledge networks of UIC in ICT industry in year 2006 and year 2017. The clusters in Fig. 4 subfigure (a) showed the ICT main topics in year 2006, which had G01N (Investigating or analyzing materials by determining their chemical or physical properties), G06F (Electric digital data processing), H04L (Transmission of digital information), H04N (Pictorial communication) and H01L (Semiconductor devices; electric solid-state devices not otherwise provided for) as core knowledge elements.

![Knowledge networks in year 2006 and year 2017](image)

**Fig. 4.** Knowledge networks in year 2006 and year 2017: subfigure (a) presents the knowledge network in year 2006; subfigure (b) shows the knowledge network in year 2017.

In year 2017, the network has evolved with much higher density with more complex characteristics. As shown in Fig. 4 subfigure (b), the topic related to G01N (Investigating or analyzing materials by determining their chemical or physical properties) drew away from other main clusters. The size of H01L (Semiconductor devices; electric solid-state devices not otherwise provided for) and related codes did not grow observably, implying that this is not the main effort of UIC in China. On the other hand, the communities of G06F (Electric digital data processing) and H04L (Transmission of digital information) grew markedly during the 12 years, which fit the technological advance under the impact of digitalization. In addition, a new cluster of knowledge interaction was developed in ICT, which led by G01R (Measuring electric variables; measuring magnetic variables).

### 4.3 UIC Multiplex network evolution analysis

After multiplex networks construction, we proceeded network analysis on annual patent data to illustrate the change of network cluster characteristics and negotiability in the process of the multiplex network evolution. The number of total nodes, featured nodes and edges are computed first to interpret topological properties of the multiplex networks. We then calculated the network density, clustering coefficient and average
path length to reveal cluster characteristics and negotiability of UIC interaction system in ICT area. Table 1 and Table 2 listed all the calculated structural characteristics in the two-layers multiplex network evolution.

Table 1. The structural characteristics in collaboration network evolution

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>194</td>
<td>359</td>
<td>499</td>
<td>608</td>
<td>499</td>
<td>891</td>
<td>1011</td>
<td>1348</td>
<td>1612</td>
<td>1831</td>
<td>1791</td>
<td>2194</td>
</tr>
<tr>
<td>I-Nodes</td>
<td>138</td>
<td>251</td>
<td>361</td>
<td>430</td>
<td>344</td>
<td>675</td>
<td>744</td>
<td>984</td>
<td>1212</td>
<td>1377</td>
<td>1352</td>
<td>1687</td>
</tr>
<tr>
<td>U-Nodes</td>
<td>56</td>
<td>108</td>
<td>138</td>
<td>178</td>
<td>155</td>
<td>216</td>
<td>267</td>
<td>364</td>
<td>400</td>
<td>454</td>
<td>439</td>
<td>507</td>
</tr>
<tr>
<td>Edges</td>
<td>437</td>
<td>780</td>
<td>1110</td>
<td>1292</td>
<td>1012</td>
<td>1865</td>
<td>2136</td>
<td>2996</td>
<td>3734</td>
<td>4302</td>
<td>4079</td>
<td>5043</td>
</tr>
<tr>
<td>Density</td>
<td>0.023</td>
<td>0.012</td>
<td>0.009</td>
<td>0.007</td>
<td>0.008</td>
<td>0.005</td>
<td>0.004</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>CC</td>
<td>0.851</td>
<td>0.798</td>
<td>0.776</td>
<td>0.773</td>
<td>0.72</td>
<td>0.697</td>
<td>0.695</td>
<td>0.701</td>
<td>0.684</td>
<td>0.698</td>
<td>0.675</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 2. The structural characteristics in knowledge network evolution

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>73</td>
<td>117</td>
<td>139</td>
<td>166</td>
<td>162</td>
<td>191</td>
<td>192</td>
<td>194</td>
<td>202</td>
<td>208</td>
<td>205</td>
<td>250</td>
</tr>
<tr>
<td>Edges</td>
<td>190</td>
<td>375</td>
<td>487</td>
<td>638</td>
<td>594</td>
<td>841</td>
<td>750</td>
<td>717</td>
<td>816</td>
<td>895</td>
<td>858</td>
<td>1198</td>
</tr>
<tr>
<td>Density</td>
<td>0.072</td>
<td>0.055</td>
<td>0.051</td>
<td>0.047</td>
<td>0.046</td>
<td>0.046</td>
<td>0.041</td>
<td>0.038</td>
<td>0.04</td>
<td>0.042</td>
<td>0.041</td>
<td>0.038</td>
</tr>
<tr>
<td>CC</td>
<td>0.554</td>
<td>0.547</td>
<td>0.554</td>
<td>0.539</td>
<td>0.573</td>
<td>0.519</td>
<td>0.508</td>
<td>0.483</td>
<td>0.499</td>
<td>0.436</td>
<td>0.501</td>
<td>0.522</td>
</tr>
<tr>
<td>APL</td>
<td>3.344</td>
<td>3.143</td>
<td>2.983</td>
<td>2.945</td>
<td>2.945</td>
<td>3.02</td>
<td>2.846</td>
<td>2.897</td>
<td>2.824</td>
<td>2.882</td>
<td>2.642</td>
<td>2.727</td>
</tr>
</tbody>
</table>

We illustrated the topological features of the multiplex network in Fig. 5. For the collaboration network layer, as shown in Fig.5 subfigure (a), the number of participating firms rose more sharply than the number of universities; the connections of the two communities grew dramatically since 2012. This trend indicated that industrial organizations in China started to actively participate UIC after year 2011. For the knowledge network, as shown in Fig.5 subfigure (b), the number of nodes gradually increasing during the 12 years, showing the ICT related technologies steadily expand with the assistance of UIC, while the interactions of these knowledge elements became increasingly frequent while the knowledge was expanding.

2 Clustering Coefficient
3 Average Path Length
Fig. 5. The trend of number of nodes, featured nodes and edges growth in the multiplex networks: subfigure (a) presents topological properties of the collaboration network evolution; subfigure (b) shows topological properties of the knowledge network evolution.

Network density describes the connectivity of a network, in which a smaller density value indicates the behaviors of network elements will receive less influence from the network structure. The global clustering coefficient measures the tendency of vertices to cluster together. Fig.6 depicted the evolution trend of network density and clustering coefficient to measure the cluster characteristics based on the normalized valued provided in Table 1 and Table 2. The network density trend of both collaboration networks (marked in light blue) and knowledge network (marked in light green) decreased by years, indicating connections between both collaboration participants and knowledge elements they have created were becoming looser. The trend clustering coefficients of collaboration networks also steady decreased, while the knowledge networks had an opposite tendency. After year 2015, the existence of groups of nodes in the knowledge network were more densely connected.

Fig. 6. Trend of clustering characteristics in multiplex network evolution

The evolution of cluster characteristics of the multiplex networks illustrated that the cooperative innovation between enterprises and academic research institutions was getting easier and wider, thus the scale of UIC was markedly broadened. There
were more potential collaborators to choose recently than before. However, for knowledge network, some of the elements started to show agglomeration advantages in gathering scientific and technological knowledge from 2015. The network had scattered and scaled agglomeration features.

We finally examined the negotiability of the UIC multiplex network for ICT industry. Fig. 7 presented the evolution trend of how negotiable of the network in two dimensions of partner collaboration and knowledge generation, based on the normalized average path length provided in Table 1 and Table 2. As illustrated via blue curve, the collaboration network of Chinese ICT area has become more negotiable from year 2011. There were less cost any two nodes needed to pay when building a connection since then. The negotiability of the knowledge work increased steadily. The network became much more efficient than before, however, the information could not reach as far as before, which implied the research and development in this area was pushed to a more refining level.

![Fig. 7. Trend of negotiability in multiplex network evolution](image)

### 4.4 Discussion

ICT-related technologies are pervasive nowadays. This empirical study illustrates that the UIC activities is characterized with sharp growth and high level of pervasiveness. There were considerable entry of new universities and firms for ICT area from year 2006 to 2017. The multiplex network we constructed contains heterogeneous information, and the evolution of collaboration and knowledge network layers presented different cluster characteristics and negotiability. With 12 years development, now it is more convenient for both universities and firms to build a connection in this area. However, the knowledge network started to show rich-club phenomenon a certain extent. Some of the knowledge elements had more agglomeration advantages than others, which will have potential influence on future collaboration activities.

This empirical study applied 4-digits IPC codes as straightforward proxies of technological knowledge considering the availability of data. Although IPC nodes can provide concise description of knowledge elements in ICT, they are not able to reveal
the detailed topics and semantics of the target area. Further semantic analytics can be conducted using patent abstracts or claims with scientific text mining techniques. In other words, the knowledge layer of the multiplex network can be replaced with a more precise version to illustrate the detailed content of research and application topics.

5 Conclusion, limitation and future work

In this short research paper, we explored a new perspective of modeling UIC interactions by building both collaboration networks and knowledge networks for future research in mechanism and opportunity discovery. Although this paper provides heuristic research on summarizing existence of collaboration network and knowledge network in UIC in to a multiplex framework, it has several limitations that need be explored in future research: (1) the knowledge elements we applied are four-digits IPC codes, which provided concise proxies of technological concepts, however, these tags are not able to present the actual content and semantics of patents; (2) this paper did not dig deeply into the dynamic mechanism of how interactions in collaboration network drive the knowledge creation and exchange in knowledge network; (3) only limited network structure indices were applied when profiling the dual-networks evolution.

From the perspective of scientific text mining, both IPC codes and keywords listed by authors or applicants are just proxies of ‘knowledge element’. Exiting research have not reached an agreement on whether these proxies are sufficient or how to extract more informative ‘elements’ to model the content of the target corpus. We will address this issue in our future research, to keep improving performance of the multiplex network in representing and analyzing the complex system of UIC.

Acknowledgements

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

References


