Network Proximity and Communities in Innovation Clusters Across Knowledge, Business, and Geography: Evidence From China

Yuan Zhou, Zhaofu Li, Yufei Liu, and Fankang Deng

Abstract—**The notion of proximity for innovation clusters needs to expand compared to traditional agglomeration literature because it involves multiple dimensions such as geography, business, and knowledge; however, limited research probed into this–it requires multidimensional data and novel methods. This article, therefore, proposes a three-layered framework that uses multisource heterogeneous data and network methods to measure the cluster-proximity in innovation clusters, in order to understand better the combined proximity between organizations within/across network communities. First, we developed a three-layered framework to map the network-based innovation cluster by integrating patent citation, business transaction, and geographic data. Second, in the innovation cluster, we identified the network communities in knowledge, business, and geographic layers, respectively. Third, we measured the cluster proximity within/across communities by using a combined index. We selected A City's machine tool sector in China as a case. This article found that machine-tool firms/organizations have higher cluster proximity within geographic communities that are enriched mostly by business connections. In comparison, they have lower cluster proximity across geographic communities, but the proximity is enhanced both by business connections and knowledge linkages. This may imply that knowledge linkages are more important in across-community proximity, and this needs more policy attention.**

Index Terms—**Across- and within-community, cluster proximity, innovation clusters, machine tool sector, multilayered heterogeneous network, multisource heterogeneous data.**

Manuscript received June 29, 2019; revised January 31, 2020 and August 31, 2020; accepted October 6, 2020. Date of publication December 1, 2020; date of current version July 2, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 71974107, Grant 91646102, and Grant 71872019, and Grant L1824039, and Grant L1724034, and Grant L1724026, the MOE (Ministry of Education in China) Project of Humanities and Social Sciences under Grant 16JDGC011, the Construction Project of China Knowledge Center for Engineering Sciences and Technology under Grant CKCEST-2020-2-5, the UK–China Industry Academia Partnership Program under Grant UK-CIAPP/260, and the Tsinghua University Project of Volvo-Supported Green Economy and Sustainable Development under Grant 20153000181. Review of this manuscript was arranged by Department Editor Y. Zhang. *(Corresponding author. Yufei Liu.)*

Yuan Zhou is with the School of Public Policy and Management, Tsinghua University, Beijing 100084, China (e-mail: zhou_yuan@mail.tsinghua.edu.cn).

Zhaofu Li and Fankang Deng are with the National Numerical Control Systems Engineering Research Center, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: lizhaofu0215@163.com; 2912105520@qq.com).

Yufei Liu is with the School of Public Policy and Management, Tsinghua University, Beijing 100084, China, and also with the Center for Strategic Studies, Chinese Academy of Engineering, Beijing 100088, China (e-mail: liuyufei0418@qq.com).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TEM.2020.3032160

I. Introduction

THE THEORY of innovation cluster goes through three phases: geographical agglomeration [1], [2], industrial clusters [3], and innovation cluster [4], [5]. Traditional literature on agglomeration or industrial cluster mostly used the notion of geographic proximity to measure the distance between firms or organizations within clusters [6], by using single-dimensional data such as geographic data or business data. However, these single-dimensional measures are unable to depict the multidimensional characteristics within innovation clusters, which involves multiple layers such as geography, business, and knowledge [7], [8]. Some recent literature attempted to use multisource data to explore the innovation clusters [9], [10]; yet, they have not extended the multisource data to specify the distance or closeness between firms/organizations within an innovation cluster. This article proposed a multidimensional concept–cluster proximity–to measure the combined distance or closeness, and this notion of cluster proximity can also leverage the prior concept of network proximity [11], [12] that uses network indicators to measure the proximity between key nodes such as firms or organizations.

Existing network literature can bring some insights to innovation cluster theory. First, they used the indicators of network proximity such as centralities that can help to better specify the closeness or distance within an innovation cluster [13]. Second, some recent network literature attempted to investigate the connections within or across network communities–this can also contribute to the measures of proximity within innovation clusters [14]. Furthermore, some recent network literature investigated multilayered heterogeneous networks, and used these concepts to study the innovation systems that involve multiple dimensions such as knowledge, business, and geography. For example, some researches attempted to study the innovation ecosystem [15] and the development of the machine tool domain [16] through multilayered heterogeneous networks by using multisource data such as patents and business [17]. Based on these literature, this article proposed that we can use multisource data to develop multilayer heterogeneous networks, and use these combined indicators to measure cluster proximity in innovation clusters.

From a methodological point of view, some recent computer science research has developed a series of new methods to analyze network, including examining the interlayers between network communities [14], developing the indicators that measure the effects of one network on the other communities [15], [16], and developing the methods that can detect the community structure of social networks [18], [19]–these methods haven't been brought into the innovation cluster literature, and these will provide new opportunities to measure cluster proximity in innovation clusters through a network lens.

This article, therefore, develops a three-layered framework that uses multisource heterogeneous data to measure cluster proximity in innovation clusters. First, we build a three-layer heterogeneous network through patent citation data, business transaction data, geographic data, in order to map the network-based innovation cluster. Second, we identify the network communities of knowledge, business, and geographic layers separately, and map core communities in different layers. Third, we measured the cluster proximity within/across communities by using a combined index, and based on which we attempt to abserve the heterogeneities of proximities. We selected A City's machine tool sector in China as a case.

The rest of this article is organized as follows. Section II reviews the literature. In Section III, the data used in this article are introduced, and we develop a framework for measuring cluster proximity in innovation clusters. In Section IV, we select A City's machine tool sector in China as a case, analyze the data and present the results. In Section V, we discuss the key findings. Finally, Section VI concludes this article.

II. Literature Review

There are two parts in this section. The first part is the review of knowledge, business, and geographic in innovation clusters. The second part reviews the methods of multilayered heterogeneous network and community detection.

A. Knowledge, Business, and Geography in Innovation Clusters

Traditional literature has studied innovation clusters by using single-dimensional data only [20], [21]. For example, Maggioni *et al.* used Web of Science literature data to explore the knowledge aspects of the innovation cluster [7]. Muro and Katz used business data to study the commercial characteristics of innovation clusters [8]. Moreno *et al.* used patent data to identify the extent to which the degree of specialization or diversity in a region may affect the innovative activities in a particular local industry [21]. Turkina *et al.* used patent citations data to study the knowledge flow within the innovation cluster [22].

In recent years, researchers have begun to use multisource data to explore the effect of other data sources in the innovation cluster [23]. For example, Tanimoto and Doi created an indicator system using geographic data and business data to examine the combined effect of geographic and business activities on innovation clusters in the San Francisco Bay area [24]. Salvador *et al.* used a questionnaire and interview data to explore the geographic characteristics of the science park as an innovation cluster [25]. Arthurs *et al.* proposed an indicator system comprising six constructs and 34 variables to explore the characteristics of innovation clusters [26]. Moreno *et al.* proposed an indicator system using patent, econometric, geographical data to explore geography and knowledge in an innovation cluster [2]. Jaffe *et al.* used patent citation data and geographic data and used metrological methods to explore knowledge and geography in innovation clusters [10].

However, these studies failed to probe into the innovation clusters through multilayered networks perspectives [15], [16], and this paved the ways to measure cluster proximity in innovation clusters using the multiple layers as different networks [17], especially when concerned with the innovation clusters that can be viewed as communities in various layers – relevant research remains sparse.

In sum, there is very limited research that examines the innovation clusters by using multilayered heterogeneous network that involves knowledge, business, and geographic layers, and none has investigated the cluster proximity in innovation clusters– this needs further research.

B. Method of Multilayered Heterogeneous Network and Network Community Detection

Innovation clusters are a multidimensional concept that can be represented using multilayered heterogeneous network. Multilayered heterogeneous network allows the representation and analysis of intralayer and interlayer connections multilayered that consist of a fixed set of nodes connected by different types of links [27].

At present, there are several methods applied to network community detection [28], and the research has shown the methods are effective. Social networks that have a topological structure of interconnected nodes that combine organization and randomness [18]. For complex large-scale networks, it is difficult to observe their characteristics directly, which requires a new method to extract comprehensive information from its structure [29]. A promising approach consists in decomposing the networks into sub-units or communities. Girvan and Newman proposed the Girvan–Newman algorithm (GN algorithm) is a topological clustering algorithm, which is suitable for the analysis and processing of network data [30]. The proposed modularity function solves the problem of over-reliance on the initial solution and the lack of global optimization goals. Newman proposed the Fast-Newman (FN) algorithm to optimize the modularity function and improve the computation, which ignores the hierarchical of the community structure [19]. Vincent D *et al.* proposed the Louvain algorithm, which is a heuristic method that is based on modularity optimization [18], which greatly reduces the computing time of community detection in complex large-scale networks. These studies supported for community detection in different layers of a multilayered heterogeneous network.

The multilayered heterogeneous network analysis methods provided support for association analysis between layers. The association analysis method is used to analyze the interaction between networks and between communities. Xu *et al.* developed a framework to investigate the interplay and innovation capacities of a multilayer innovation ecosystem that involved science, technology, and business [15]. Murata proposed a layer-to-layer coupling analysis method for multilayer social networks and attempted to use multiple interlayer connections to detect communities in multilayer social networks [31]. Cheng *et al.* proposed an effective index to measure the effect of multilayer network interactions [32]. These studies will provide new opportunities to measure cluster proximity based on multisource heterogeneous data in innovation clusters through a network lens, especially on exploring the increasing knowledge and business networks on traditional geographic communities.

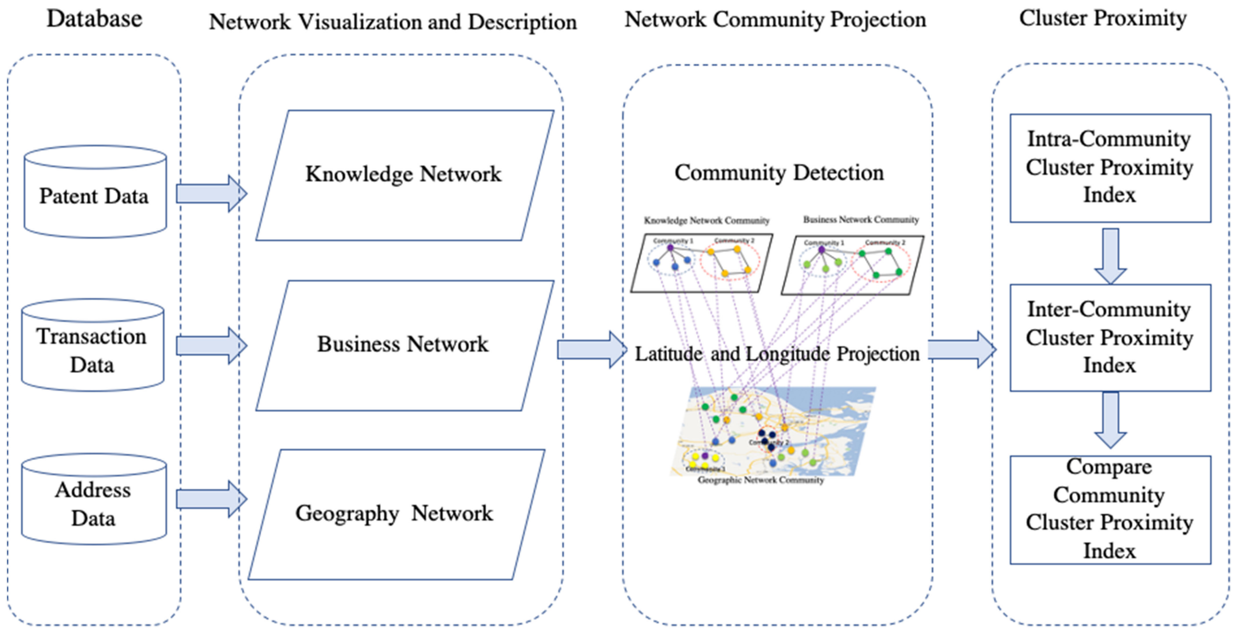


Fig. 1. Research framework.

III. Methodology and Data

The novel framework proposed herein integrates multilayered heterogeneous network analysis, the Louvain clustering algorithm, cluster proximity measurement, and the visualization methods to measure cluster proximity based on multisource heterogeneous data in innovation clusters.

A. Framework

A framework, based on multilayered heterogeneous network that includes knowledge, business, and geographic layers, is proposed to measure cluster proximity in innovation clusters.

First, we build a three-layer heterogeneous through three-dimensional data–knowledge network is built from patent data, business network is built from transaction data, and geographic network is built from geographic data. Second, communities are obtained by clustering three networks, respectively, and we map core communities obtained from the knowledge network and the business network to the geographic network. Third, we use a combined proximity index with/access communities to measure the cluster proximity in innovation clusters when the knowledge network and the business network are introduced.

This framework is an integration of many methods as shown in Fig. 1.

Besides, on the one hand, we need some network measurement indicators to measure the single-layer network, which is necessary when describing the differences between knowledge, business, and geographic network. On the other hand, we need some visualization tools to help us better understand the multilayer network structure. These visualization tools include:NetworkX,[1] Echarts,[2] Pymnet,[3] and MapV Editor.[4]

[1][Online]. Available: http://networkx.github.io
[2][Online]. Available: https://www.echartsjs.com/zh/index.html
[3][Online]. Available: http://www.mkivela.com/pymnet/tutorial.html
[4][Online]. Available: https://mapv.baidu.com/editor/

B. Data Collection

This article collected three datasets: knowledge, business, and geography. We use patent direct citation relationship to build the knowledge network in the innovation cluster [10]. Patent data were used to analyze the knowledge creation and flow within industries and across national borders [33], [34]. The Derwent Innovation database[5] is used as the data source [35], [36] and the Chinese State Intellectual Property Office database is used as a Chinese patent data source for Chinese and English matching [37]. With enterprises as nodes, citations were used as ties to build a knowledge network. Transaction data directly reflects the money transactions between enterprises and can be used to analyze business systems. We collected a large amount of transaction data as a data source. With enterprises as nodes, money transactions were used as ties to build a business network. Address data represents the address location of an enterprise and can be used to analyze geographic proximity. We collected the addresses of a series of enterprises as a data source, and each enterprise had an address in this source. With enterprises as nodes, distances were used as ties to build a geographic network.

C. Network Construction, Measurement, and Visulization

We can build a multilayer network based on the collected data. In order to compare network topology of different networks, we introduce a series of single-layer network measurement indicators to describe the single-layer network. These indicators are shown in Table I.

To observe the network more intuitively, we developed a network visualization tool. The tool includes single-layer social network visualization, multilayer social network visualization, and community detection and projection on a map. The single-layer social network visualization tool is based on NetworkX and Echarts, which can be used to perform operations and data analysis on the network and visualize it. The visual layout is

[5][Online]. Available: https://www.derwentinnovation.com

TABLE I
INDICATORS PROPOSED TO MEASURE THE NETWORK AND NODES

Indicators	Indicators meaning
Density	Number of lines in a graph, expressed as a proportion of the maximum possible number of lines [38]
Clustering coefficient	Measure of the extent to which nodes tend to cluster together in a graph [39]
Shortest path length	Path of minimum total length between two given nodes s and t in a network [40]

based on the force-guided layout. NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. The multilayer social network visualization tool is based on Pymnet, which can be used to perform operations and data analysis on a network and visualize it. Pymnet is a free library for analyzing multilayer social networks. It creates visualizations using Matplotlib or D3 as the backend.

D. Community Detection and Visualization

Since we believe that in innovation clusters, cluster proximity is different within/across communities, so the networks need to be divided into communities first. Community detection can divide the network into communities. The nodes in the same community are densely connected, while the nodes in different communities are sparse connected [18]. Louvain algorithm is an effective community detection method, which has the advantages of fast speed and high accuracy. The modularity function is originally used to measure the quality of the community discovery algorithm results and to characterize the closeness of the communities found [19]. Since it can describe the intimacy of the community, it can be used as an optimization function, which is defined as follows:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta\left(c_i, c_j\right) \tag{1}$$

where A_{ij} represents the weight of the edge between i and j, $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to vertex i, c_i is the community to which vertex i is assigned, the δ-function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise and $m = \frac{1}{2} \sum_{ij} A_{ij}$. This algorithm provides a fair compromise between the estimation accuracy of the modular maximum.

To observe the community structure more clearly, we developed a community visualization tool. This tool integrates the Louvain algorithm, which can perform community detection for large-scale complex networks of which a human can not easily observe the morphology and features with the naked eye. We used the MapV Editor to visualize the main community projection on a map.

E. Cluster Proximity Index Measurement

Cluster proximity is used to measure the proximity between enterprises in innovation clusters. Existing research define cluster proximity as the geographic distance between the focal firm and its suppliers [6]. The definition of cluster proximity is based on geographical agglomeration in the existing research. In the innovation cluster, not only the geographical layer is concerned, but also the business and knowledge layers are concerned. In geographic, business, and knowledge networks, the shortest path length between two enterprises in the network is used to represent geographic proximity, business proximity and knowledge proximity. Therefore, we define cluster proximity as the superposition of knowledge proximity, business proximity and geographic proximity.

Compared with a single-layer network, a multilayer network can introduce more connections for nodes, which can shorten the shortest distance between nodes and reduce cluster proximity. We introduce cluster proximity index for each community to descript the degree of cluster proximity reduction when the system joins a new network. Obviously, the greater of a community's the cluster proximity index, the more stronger the role by the newly added network on the community. The cluster proximity index are divided into intra-community cluster proximity index and inter-community cluster proximity index.

1) Cluster Proximity Index Within Communities: Cluster proximity index within communities H_{intra_α} is defined to quantify to what extent of the cluster proximity in network B community α is decreased with network A, which is defined as follows:

$$H_{intra\_\alpha} = \sum_{\substack{n_1 = 1 \\ n_1 \neq n_2}}^{n_\alpha} \left[d^{(1)}_{n_1,\, n_2} - d^{(2)}_{n_1,\, n_2} \right] \tag{2}$$

where $d^{(1)}_{n_1,\, n_2}$ is the shortest path length from node n_1 to n_2 without network A, and $d^{(2)}_{n_1,\, n_2}$ is that with network A.

2) Cluster Proximity Index Across Communities: Cluster proximity index across communities index H_{inter_α} is defined to quantify to what extent of the cluster proximity in network B community α is decreased with network A [32], which is defined as follow:

$$H_{inter\_\alpha} = \sum_{\substack{\beta = 1 \\ \beta \neq \alpha}}^{c} \left[d^{(1)}_{\alpha,\, \beta} - d^{(2)}_{\alpha,\, \beta} \right] \tag{3}$$

where $d^{(1)}_{\alpha,\, \beta}$ is the distance from community α to β, i.e., averaged pairwise distances from the nodes in community α to those in community β without network A, and $d^{(2)}_{\alpha,\, \beta}$ is that with network A.

IV. RESULTS

In the previous sections, we have proposed a framework to measure the cluster proximity in innovation cluster. In this section, we seclect A City's machine tool sector in China as a case.

A. Data Collection

The machine tool sector is representative in terms of innovation clusters. Machine tools in the machine tool sector are indispensable instruments in the manufacturing sector, and they are needed in almost all machining processes, so the machine tool sector have always occupied an important position in geographic and business. With the improvement of product

(a) (b) (c)

Fig. 2. Single-layer network visualization: (a) knowledge network, (b) business network, and (c) geographic network.

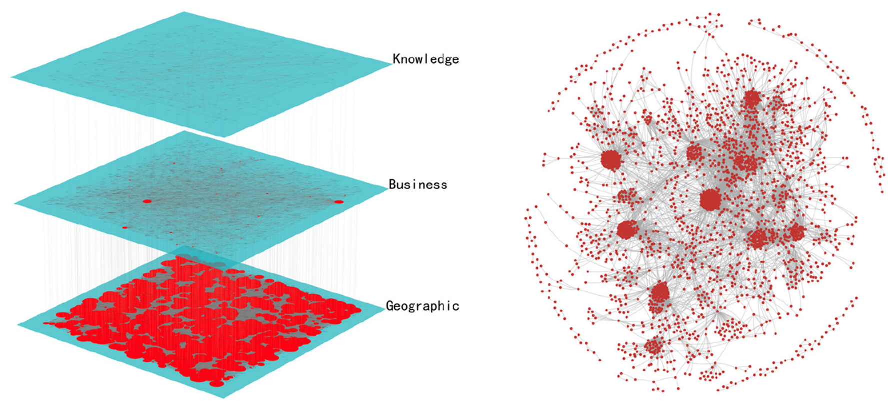


Fig. 3. (a) Multilayer social network. (b) Aggregate network visualization.

processing quality, there is an urgent need for the innovation of the machine tool sector. Machine tools play a unique and critical role in the sector [41], and machine tool technology is likely to become a pillar technology of the new generation of manufacturing [10]. Manufacturing is the pillar sector of A City's economy. The city took the lead in establishing a modern manufacturing system with complete categories and became an important advanced manufacturing base in China.

We collected three datasets: knowledge, business, and geography. We retrieved worldwide patent data from the DWPI and DPCI databases in Derwent Innovation as the seeds for building a knowledge network. The search queries were "IC = ((B23) OR (G05B001918)) AND AD > = (19850101) AND AD < = (20171231) AND PRC = (CN)". The machine tool field includes machine tools and CNC systems, where IC represents international patent classification (IPC), B23 represents the machine tool, and G05B001918 represents the CNC system. AD represents the application date, PRC represents the priority country, and CN represents China. In total, 388,725 patent applications were retrieved from the database. We extracted patent data for A City, and 8372 patents were selected. We retrieved worldwide patent data from the Chinese State Intellectual Property Office database for Chinese and English matching of enterprise names. The search queries were "select ∗ from cnipr_patent.cnipr_cn_all where 'main classification number 'LIKE '%B23%' or 'main classification number ' LIKE '%G05B19/18%'" from the SQL database. The search period is set to 1985–2017. In total, 472 927 patent applications were retrieved from the database. The transaction data of the machine tool sector is collected during 2016-2017 from A City, China and the number of transaction data records is 846 588. Our dataset contained the following four fields: transaction content, transaction money, transaction specification, and transaction unit. We collected the addresses of a series of enterprises during 2016-2017 for building a geographic network. In total, information of 1 463 674 enterprises is collected. We then extracted the address data of A City enterprises, and 109 752 enterprises were selected. We calculated the latitudes and longitudes of these enterprises through the API of the Baidu map and then calculated the distance between the two enterprises based on latitudes and longitudes.

B. *Multilayered Heterogeneous Network Construction*

A multilayered heterogeneous network includes knowledge, business, and geographic layers. First, we built knowledge, business and geographic networks. Second, we built a multilayered heterogeneous network. Third, we analyzed network topology of knowledge, business, geographic, and aggregate networks.

1) Knowledge, Business, and Geographic Networks: We used patent citation data, business transaction data, and geographic data to build knowledge, business and geographic networks. To build a knowledge network, we extracted patent applications and removed the names of the person on each application. With enterprises as nodes, citations were used as ties to build a knowledge network $G_{Knowledge}$ as shown in Fig. 2(a). To build a business network, we extracted all transaction data, and each transaction data point is a transaction between two enterprises. With enterprises as nodes, money transactions were used as ties to build a business network. This network is a transaction network between enterprises of A City's machine tool sector and nationwide enterprises. To analyze the business network of the machine tool sector in A City, we have extracted all the enterprises in A City's machine tool sector and built a sub-network of this network $G_{Business}$ as shown in Fig. 2(b). To build a geographic network, we extracted the addresses of A City's machine tool sector from the address data of all enterprises. We then calculated the distance between any two enterprises. When the distance between them is less than 3 km, we considered the two enterprises to be geographically adjacent. With enterprises as nodes, distances were used as ties to build a geographic network $G_{Geographic}$ as shown in Fig. 2(c).

As shown in Fig. 2, the knowledge network consisted of 268 nodes and 226 edges, and its density is 0.0063, and the business network consisted of 1446 nodes and 1573 edges, and its density is 0.0015, and the geographic network consisted of 1593 nodes and 45 611 edges, and its density is 0.0360. The knowledge network had not significant community, and the business and geographic networks had significant community.

2) Multilayered Heterogeneous Network: We build a multilayer heterogeneous network M_1 that includes knowledge, business, and geographic layers. We use knowledge network $G_{\text{Knowledge}}$, business network G_{Business}, and geographic network $G_{\text{Geographic}}$ to build a multilayered heterogeneous network, which use the enterprises as the nodes and the different types of node connections as the edges. The common enterprises in different layers also are connected. Multilayer heterogeneous network is shown in Fig. 3(a). The number of the identical nodes in the knowledge and geographic layers is less than the number of the identical nodes in the business and geographic layers, and the number of nodes and edges were very different at the knowledge, business, and geographic layers.

Then, we aggregated the knowledge, business, and geographic layers into an aggregate network $G_{\text{Aggregate}}$, as shown in Fig. 3(b). As shown in Fig. 3(b), the aggregate network consisted of 2548 nodes and 47 338 edges, and its density is 0.0146, and the aggregate network had significant community. As shown in Figs. 3(b) and 2(c), the aggregate network and geographic network were similar.

3) Network Topology of Knowledge, Business, Geographic, and Aggregate Networks: We used a series of indicators to measure knowledge, business, geographic, and aggregate networks for analyzing their topology. To analyze the network topology of the knowledge network $G_{\text{Knowledge}}$, the business network G_{Business}, and the geographic network $G_{\text{Geographic}}$, we generated a randomized degree-preserving counterpart of the previous network for comparison, denoted as network $G_{\text{Random-Knowledge}}$, $G_{\text{Random-Business}}$, $G_{\text{Random-Geographic}}$ and $G_{\text{Random-Aggregate}}$. The fundamental descriptive statistics of these network topologies is shown in Table II. n is the number of nodes, e is the number of edges, d is the density of the network, c is the clustering coefficient, and p is averaged shortest path length.

We got the network topology of knowledge, business, geographic, and aggregate networks by comparing the statistical description in Table II. As shown in Table II, the averaged shortest path distance of network $G_{\text{Aggregate}}$ is 4.2561, shorter than that of network $G_{\text{Geographic}}$, meaning that the innovation cluster benefits from business and knowledge networks in A City. The averaged shortest path length of the knowledge network $G_{\text{Knowledge}}$ is significantly lower than $G_{\text{Random-Knowledge}}$, and the clustering coefficient of the network $G_{\text{Knowledge}}$ is significantly larger than $G_{\text{Random-Knowledge}}$. The above two points show that the knowledge network is not random, and it is typically a small-world network. Each of these enterprises has many opportunities, and there are no monopoly enterprises. Similarly, comparing G_{Business} and $G_{\text{Random-Business}}$, $G_{\text{Geographic}}$ and $G_{\text{Random-Geographic}}$, and $G_{\text{Aggregate}}$ and $G_{\text{Random\_Aggregate}}$, it is evident that the business network G_{Business} is not random, and it is typically a small-world network, Futhermore, the geographic network $G_{\text{Geographic}}$ and the aggregate network $G_{\text{Aggregate}}$ is not random, not a regular network and not a small world network. Compared to a random network, the information transmission in the small-world network is faster, and the structure of the small-world network will inevitably have many communities and node groups [33].

TABLE II
STATISTICAL DESCRIPTION OF THE CONSTRUCTED NETWORKS

ID	n	e	d	c	p
$G_{Knowledge}$	268	226	0.0063	0.0974	4.8529
$G_{Random-Knowledge}$	268	226	0.0063	0.0103	10.3274
$G_{Business}$	1446	1573	0.0015	0.0149	6.1136
$G_{Random-Business}$	1446	1586	0.0015	0.0004	8.6591
$G_{Geographic}$	1593	45611	0.0360	0.8527	7.1042
$G_{Random-Geographic}$	1593	45611	0.0360	0.0356	2.0867
$G_{Aggregate}$	2548	47338	0.0146	0.5063	4.2561
$G_{Random-Aggregate}$	2548	47338	0.0146	0.0142	2.5584

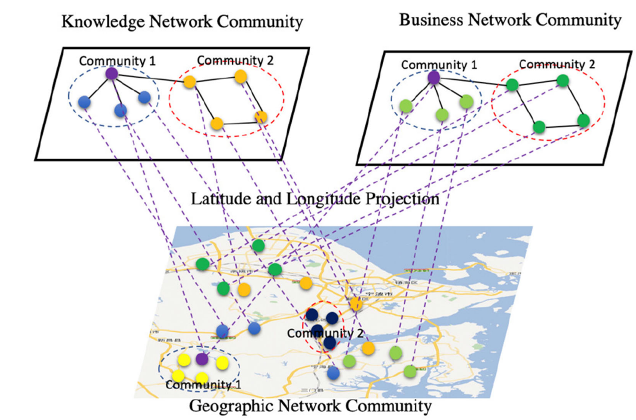
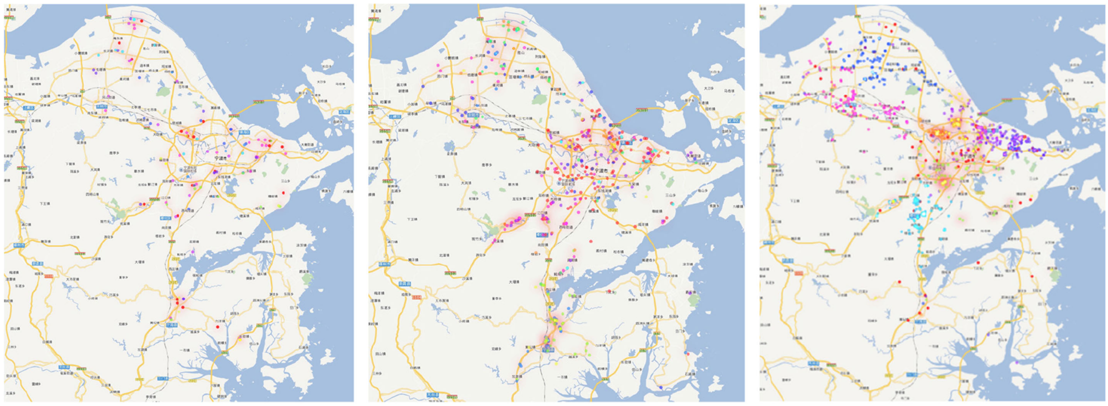


Fig. 4. Example of network communities projection on a map.

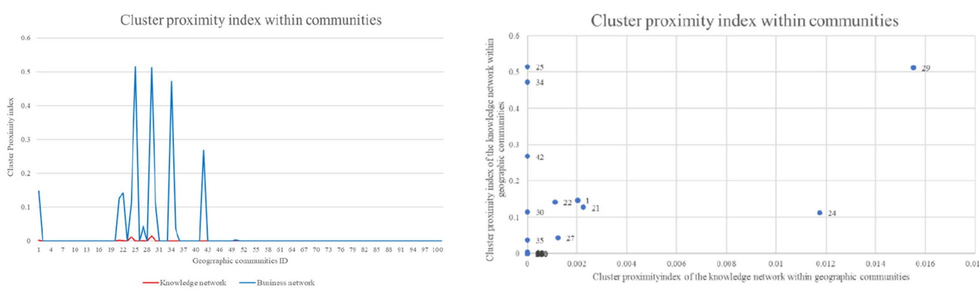
C. Projection of Knowledge and Business Network Communities on Map and Comparison With Geographic Communities

Social networks that have a topological structure of interconnected nodes that combine organization and randomness [18]. Therefore, cluster proximity in different regions is different. To discovery the cluster proximity different in different regions,



(a) (b) (c)

Fig. 5. Top ten communities projection on the map: (a) Knowledge network, (b) business network, and (c) geographic network.



(a) (b)

Fig. 6. (a) Cluster proximity index within communities. (b) Cluster proximity index within communities of knowledge network versus business network.

we then identified the communities of knowledge, business, and geographic layers, respectively, and compared the differences of core communities projection on a map between these layers, the example of network communities projection on a map as shown in Fig. 4.

We identified the community structures in the network using the Louvain algorithm. The knowledge network partitioned into 65 nonoverlapping communities. The business network partitioned into 203 nonoverlapping communities. The geographic network partitioned into 101 nonoverlapping communities. A community in the network is a set of nodes that are densely interconnected but loosely connected with the rest of the network. We sorted the community based on the number of enterprises included in the community and projected the top ten communities on the map using MapV Editor. The top ten community projection of the knowledge, business, and geographic networks are shown in Fig. 5(a)–(c), respectively, and visualization of community detection results for corresponding networks are shown in Fig. 8(a)–(c) in the Appendix.

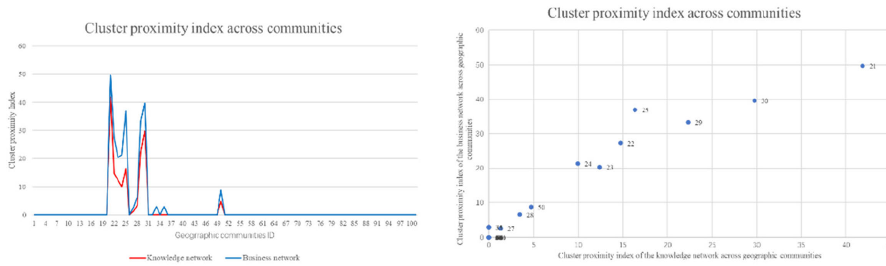
As shown in Fig. 5, the core communities of the knowledge and business networks were projected on the map [Fig. 5(a) and (b)], and we found that their core community projections were not highly overlapping with the geographic agglomeration communities [Fig. 5(c)]. This difference led to smaller cluster proximity between geographic communities with business and knowledge networks than the original geographic network.

D. Explore the Cluster Proximity Index of Geographic Communities

To explore the heterogeneous promotion role of knowledge network and business network on the cluster proximity within geographic communities and across geographic communities, respectively.

1) Explore the Cluster Proximity Index Within Geographic Communities: We calculated the promotion role of knowledge network and business network on cluster proximity within geographic communities, and the overall results are shown in Fig. 6.

We compared the cluster proximity index withincommunity to analyze the difference promotion role of knowledge network and business network on the cluster proximity within geographic communities. As shown in Fig. 6(a), compared to the two lines, the promotion role of the business network on the cluster proximity within geographic communities is bigger than that of the knowledge network, indicating that the cluster proximity within geographic communities depend more on the business network compared to the knowledge network. As shown in Fig. 6(b), the promotion role of the knowledge network on the cluster proximity index within geographic communities is less than 0.016, and the promotion role of the business network on the cluster proximity index within geographic communities is less than 0.6, indicating that knowledge and business networks have a limited promotion role on the cluster proximity within



(a) (b)

Fig. 7. (a) Cluster proximity index across communities. (b) Cluster proximity index across communities of knowledge network versus business network.

geographic communities. Furthermore, the knowledge network plays a negligible role on the cluster proximity within the geographical community.

2) Explore the Across-Community Cluster Proximity Index Between Geographic Communities: To study the promotion role of knowledge and business networks on the cluster proximity across geographic communities, we calculated the promotion role of knowledge network and business network on the cluster proximity across geographic communities, and the overall results are shown in Fig. 7.

We compared the cluster proximity index across communities to analyze the difference promotion role of knowledge network and business network on the cluster proximity across geographic communities. As shown in Fig. 7(a), compared to the two lines, they were close. As shown in Fig. 7(b), the promotion role of the business network and knowledge network on the cluster proximity across geographic communities is consistent across trends. The above two points indicated that the cluster proximity across key geographic communities depended both on business network and knowledge network. As shown in Fig. 7(b), the cluster proximity index across communities of some geographic communities is bigger than ten, indicating that the knowledge network and business network clearly strengthen the cluster proximity across the geographic network communities.

V. Discussion

The application of our framework provides a holistic view of the multilayered innovation cluster of A City's machine tool sector. Both densities of the whole knowledge and business networks are low, while the densities of the whole geographic network is high, and the community structure in the business network is stronger than that in the knowledge network. Knowledge and business play a promotion role on the cluster proximity in the innovation cluster. This phenomenon comports with the evolution stage of innovation clusters [3].

Futher, we probe into the communities of knowledge, business, and geographic networks in the innovation cluster. From the community perspective, we find that the similarity of communities between knowledge, business and geographic networks are very low, which indicates that enterprises have different partners in technological citation, business transactions, and geographical location. Different enterprise types in the network composition reflected different network characteristics [42]. The difference in the composition of the three-layered network nodes in the innovation cluster of A City has led to the different structure of their communities.

Knowledge and business networks have different promotions on the cluster proximity of innovation clusters. Knowledge and business networks have different promotions on the cluster proximity within/across geographic communities. The knowledge network has a strong promotion on the cluster proximity within geographic communities, as well as the business network has a strong promotion on the cluster proximity across geographic communities. The business network has a small promotion on the cluster proximity within geographic communities, while the promotion of the knowledge network on the cluster proximity within geographic communities is negligible. The direct factor is the number of nodes and edges in the community, and the importance of the nodes in the network is also an important factor [43]. The difference of the community cluster proximity index can be affected by external factors such as the government's concentrated policies and the region's culture and institutions [42], [44].

VI. Conclusion

This article proposed a three-layered framework that uses patent citation, business transaction, geographic data, and network methods to measure the cluster-proximity in innovation clusters. The three-layered network of geography, business, and knowledge comprehensively displayed the landscape of innovation cluster. Based on the three-layered network and identification of the network communities in knowledge, business, and geographic layers, respectively, we understood better the combined proximity between organizations within/across communities. A case study of A City's machine tool sector demonstrated that machine-tool firms/organizations have higher cluster proximity within geographic communities and lower cluster proximity across geographic communities. The proximity within geographic communities is enriched mostly by business connections, but the proximity across geographic communities is enhanced both by business connections and knowledge linkages. This may imply that knowledge linkages are more important in across-community proximity, and this needs more policy attention. This article contributes to expand the proximity notion for innovation clusters. However, we consider the cluster proximity in only one city, while the innovation cluster is also affected by other cities or regions. In the future, we plan to extend the dataset to explore the effects of knowledge, business, geographic communities in innovation cluster.

Appendix

(a)The knowledge network. (b)The business network. (c)The geographical network

Fig. 8. Visualization of community detection. (a) Knowledge network. (b) Business network. (c) Geographical network.

Acknowledgment

The authors would like to thank the School of Economics and Management and the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications for their valuable collaboration in this article.

References

[1] K. Börner *et al.*, "Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy," *Proc. Natl. Acad. Sci.*, vol. 115, no. 50, pp. 12630–12637, 2018.

[2] R. Moreno, R. Paci, and S. Usai, "Geographical and sectoral clusters of innovation in europe," *Ann. Reg. Sci.*, vol. 39, no. 4, pp. 715–739, 2005.

[3] C. Xiao-hong, Z. Yuan, X. Guan-nan, and S. JUN, "Guangdong real faith science park: An example of government-driven industry transformation towards Hi-tech," *Chinese J. Manag. Sci.*, vol. 21, no. S2, pp. 751–757, 2013.

[4] J. S. Engel and I. del-Palacio, "Global networks of clusters of innovation: Accelerating the innovation process," *Bus. Horiz.*, vol. 52, no. 5, pp. 493–503, 2009.

[5] A. S. Basyuk, A. Y. Anisimov, V. V. Prokhorova, O. N. Kolomyts, and F. V. Shutilov, "Administration management in the innovation cluster," *Int. Rev. Manag. Mark.*, vol. 6, no. 6, pp. 180–184, 2016.

[6] R. Ganesh Pillai and V. Bindroo, "Supplier cluster characteristics and innovation outcomes," *J. Bus. Res.*, vol. 112, pp. 576–583, 2020.

[7] M. A. Maggioni, F. Gambarotto, and T. E. Uberti, *Mapping the Evolution of "Clusters': A Meta-Analysis*. USA: California: Berkeley Electronic Press Services, 2009.

[8] M. Muro and B. Katz, The new "cluster moment": How regional innovation clusters can foster the next economy. *Entrepreneurship and Global Competitiveness in Regional Economies: Determinants and Policy Implications*, p. 93, 2011.

[9] S. Herliana, "Regional innovation cluster for small and medium enterprises (SME): A triple helix concept," *Procedia - Soc. Behav. Sci.*, vol. 169, no. Aug. 2014, pp. 151–160, 2015.

[10] A. B. Jaffe, M. Trajtenberg, and R. Henderson, "Geographic localization of knowledge spillovers as evidenced by patent citations," *Q. J. Econ.*, vol. 108, no. 3, pp. 577–598, 1993.

[11] M. Castro, P. Druschel, Y. C. Hu, and A. Rowstron, "Exploiting network proximity in peer-to-peer overlay networks," *Citeseer*, vol. 36, no. 2, pp. 1–15, 2002. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.71.7902&rep=rep1&type=pdf

[12] R. E. Rice and A. Carolyn, "Attitudes toward new organizational technology: Network proximity as a mechanism for social information processing," *Adm. Sci. Q.*, vol. 36, no. 2, pp. 219–244, 1991.

[13] X. Hongming, W. Xianbiao, W. Su, and Y. Wei, "Research on cluster, network, and the innovation of IJVs," *Sci. Res. Manag.*, vol. V29, no. 6, pp. 23–29, 2008.

[14] T. Ahmad, M. Haque, and A. M. Khan, "*An Energy-Efficient Cluster Head Selection Using Artificial Bees Colony Optimization for Wireless Sensor Networks BT - Advances in Nature-Inspired Computing and Applications*," S. K. Shandilya, S. Shandilya, and A. K. Nagar, Eds. Berlin, Germany: : Springer International Publishing, 2019, pp. 189–203.

[15] G. Xu, Y. Wu, T. Minshall, and Y. Zhou, "Exploring innovation ecosystems across science, technology, and business: A case of 3D printing in china," *Technol. Forecast. Soc. Change*, vol. 136, pp. 208–221, 2018.

[16] J. Chen *et al.*, "Exploring the development of research, technology and business of machine tool domain in new-generation information technology environment based on machine learning," *Sustainability*, vol. 11, no. 12, 2019, Art. no. 3316.

[17] B. Oselio S. Liu, and A. Hero, *Cooperative and Graph Signal Processing: Principles and Applications*. U.K.: Mara Conner: Academic Press, 2018, ch. 25, pp. 679–697.

[18] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech. Theory Exp.*, vol. 2008, no. 10, pp. 1–12, 2008.

[19] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, 2004, Art. no. 066133.

[20] V. V. Mazur, K. A. Barmuta, S. S. Demin, E. A. Tikhomirov, and M. A. Bykovskiy, "Innovation clusters: Advantages and disadvantages," *Int. J. Econ. Financial Issues*, vol. 6, no. 1S, pp. 270–274, 2016.

[21] R. Moreno, R. Paci, and S. Usai, "Innovation clusters in the european regions," *Eur. Plan. Stud.*, vol. 14, no. 9, pp. 1235–1263, 2006.

[22] E. Turkina, B. Oreshkin, and R. Kali, "Regional innovation clusters and firm innovation performance: An interactionist approach," *Reg. Stud.*, vol. 53, no. 8, pp. 1193–1206, 2019.

[23] Y. Jung, E. Kim, and W. Kim, "The scientific and technological interdisciplinary research of government research institutes: Network analysis of the innovation cluster in South Korea," *Policy Stud.*, 2019. doi: 10.1080/01442872.2019.1593343.

[24] K. Tanimoto and M. Doi, "Social innovation cluster in action: A case study of the san francisco bay area," *Hitotsubashi J. Commer. Manage.*, vol. 41, no. 1, pp. 1–17, 2007.

[25] E. Salvador, I. Mariotti, and F. Conicella, "Science park or innovation cluster?: Similarities and differences in physical and virtual firms' agglomeration phenomena," *Int. J. Entrepreneurial Behav. Res.*, vol. 19, no. 6, pp. 656–674, 2013.

[26] D. Arthurs, E. Cassidy, C. H. Davis, and D. Wolfe, "Indicators to support innovation cluster policy," *Int. J. Technol. Manage.*, vol. 46, no. 3–4, pp. 263–279, 2009.

[27] S. E. Smith-Aguilar, F. Aureli, L. Busia, C. Schaffner, and G. Ramos-Fernández, "Using multiplex networks to capture the multidimensional nature of social structure," *Primates*, vol. 60, no. 3, pp. 277–295, 2019.

[28] Y. G. Kim, J. H. Suh, and S. C. Park, "Visualization of patent analysis for emerging technology," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1804–1812, 2008.

[29] Y. Zhou, F. Dong, D. Kong, and Y. Liu, "Unfolding the convergence process of scientific knowledge for the early identification of emerging technologies," *Technol. Forecast Soc. Change*, vol. 144, pp. 205–220, 2019.

[30] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proc. Natl. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2002.

[31] T. Murata, “Comparison of inter-layer couplings of multilayer networks,” in *Proc. - 11th Int. Conf. Signal-Image Technol. Int.-Based Syst. SITIS*, 2016, pp. 448–452.

[32] H. M. Cheng, Y. Z. Ning, X. Ma, X. Liu, and Z. Y. Zhang, “Effectiveness of rapid rail transit system in Beijing,” *PLoS One*, vol. 12, no. 7, p. e0180075, 2017, doi: 10.1371/journal.pone.0180075.

[33] R. Bekkers and A. Martinelli, “Knowledge positions in high-tech markets: Trajectories, standards, strategies and true innovators,” *Technol. Forecast. Soc. Change*, vol. 79, no. 7, pp. 1192–1216, 2012.

[34] Y. Zhou, M. Pan, and F. Urban, “Comparing the international knowledge flow of china’s wind and solar photovoltaic (PV) industries: Patent analysis and implications for sustainable development,” *Sustain.*, vol. 10, no. 6, p. 1883, 2018.

[35] J. Nordensvard, Y. Zhou, and X. Zhang, “Innovation core, innovation semi-periphery and technology transfer: The case of wind energy patents,” *Energy Policy*, vol. 120, pp. 213–227, 2018.

[36] B. Wang, Y. Liu, Y. Zhou, and Z. Wen, “Emerging nanogenerator technology in china: A review and forecast using integrating bibliometrics, patent analysis and technology roadmapping methods,” *Nano Energy*, vol. 46, pp. 322–330, 2018.

[37] M. Pan, Y. Zhou, and D. K. Zhou, “Comparing the innovation strategies of chinese and european wind turbine firms through a patent lens,” *Environ. Innov. Soc. Transitions*, vol. 30, pp. 6–18, 2019.

[38] J. Scott, “Social network analysis,” *Sociology*, vol. 22, no. 1, pp. 109–127, 1988.

[39] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertesz, “Generalizations of the clustering coefficient to weighted complex networks,” *Phys. Rev. E*, vol. 75, no. 2, 2007, Art. no. 27105.

[40] E. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numer. Math.*, vol. 1, no. 1, pp. 269–271, 1959.

[41] C. Liu, H. Vengayil, R. Y. Zhong, and X. Xu, “A systematic development method for cyber-physical machine tools,” *J. Manuf. Syst.*, vol. 48, pp. 13–24, 2018.

[42] S. Wang, Y. Shao, and B. Pang, “A study of high-end equipment manufacturing industry cluster network and innovation capability,” in *Proc. Int. Academic Conf. Frontiers Social Sci. Manage. Innov. (IAFSM 2018)*, Atlantis Press, 2019.

[43] H. Li, G. C. de Zubielqui, and A. O’Connor, “Entrepreneurial networking capacity of cluster firms: A social network perspective on how shared resources enhance firm performance,” *Small Bus. Econ.*, vol. 45, no. 3, pp. 523–541, 2015.

[44] H. D. Kim, D. H. Lee, H. Choe, and I. W. Seo, “The evolution of cluster network structure and firm growth: A study of industrial software clusters,” *Scientometrics*, vol. 99, no. 1, pp. 77–95, 2014.