Partner Selection in China Interorganizational Patent Cooperation Network Based on Link Prediction Approaches

Wei Chen 1, Hui Qu 1,∗ and Kuo Chi 2

1 School of Economics and Management, Harbin Engineering University, Harbin 150001, China; wei_chen666@hrbeu.edu.cn
2 School of Information and Communication Engineering, Hainan University, Haikou 570228, China; chikuo@hainanu.edu.cn or 994349@hainanu.edu.cn
* Correspondence: quhui@hrbeu.edu.cn; Tel.: +86-158-4662-6497

Abstract: To enhance competitiveness and protect interest, an increasing number of organizations cooperate on patent applications. Partner selection has attracted much more attention because it directly affects the success of patent cooperation. By collecting some cooperative patents applied for by different categories of organizations in China from 2007 to 2015, an interorganizational patent cooperation network was built for this paper. After analyzing certain basic properties of the network, it was found that the network possessed some typical characteristics of social networks. Moreover, the network could be divided into communities, and three communities were selected to analyze as representative. Furthermore, to explore the partner selection in the patent cooperation network, eight link prediction approaches commonly used in social networks were introduced to run on another interorganizational patent cooperation network built by the patents applied for in 2016. The precision metric results of the eight link prediction approaches show that they are effective in partnership prediction; in particular, the Common Neighbors (CN) index can be effectively applied to the selection of unfamiliar partners for organizations in patent cooperation. Moreover, this paper also verified the trust transitivity based not only on historical cooperation but also on geographical location, and the complementarity of capabilities still plays an important role in partner selection for organizations.

Keywords: partner selection; China; interorganizational patent cooperation network; link prediction

1. Introduction

With the increasing competition in business, conflicts of interest arising from intellectual property rights are becoming even more intense. As a legal protection measure, patent application has been an important way for organizations or individuals to protect their innovations and interests. In order to improve the efficiency of patent application, cooperation has become a mainstream trend that is adopted by an increasing number of organizations [1]. In patent cooperation, how to select suitable partners is crucial to innovation implementation and profit maximization, which also plays a decisive role in the success of patent application [2,3]. Literally, partner selection has become a strategy decision for cooperative patent applicants in many fields [4,5].

In practice, most patent applicants prefer to select partners they have worked with in the past than new partners even with strong capabilities [6,7]. For one thing, cooperating with old partners can reduce risks and improve the success rate of cooperative patent application. For another, it is conductive to the distribution of benefits based on previous cooperation. However, historical information may be unavailable when new partners need to be selected. In order to address the problem, many studies have been carried out to explore the partner selection from multiple perspectives [8–11], such as geographical locations of cooperators and the distance between them [12], the relevance of cooperators’ capabilities [13,14], and trust relationships between cooperators [15,16]. Most of those...
studies analyzed the influence of different perspectives on partner selection, but they lacked a generic solution to the problem.

To find a generic method to deal with partner selection for organizations in patent cooperation, we collected the data on cooperative patents applied for by organizations in China from 2007 to 2015 as the research object of this paper. The data were composed of 46,492 cooperative patents and could be represented as a network to show the different organizations and the partnerships between them for convenience. By analyzing certain basic properties of the patent cooperation network, including network diameter, average degree of nodes, average path length, and centralities of nodes, we found that this network had some typical characteristics of social networks, such as Pareto’s principle, “small-world”, and obvious community structure. Because the scale of the network was too large to be analyzed, we chose three communities as the representatives after the community detection for analysis. To further determine whether the network could be regarded as a social network, we introduced link prediction approaches to predict the possible future partnerships between different organizations in the patent cooperation network and compare the prediction results with the data of cooperative patents applied for in 2016 to verify the feasibility.

The reminder of this paper is organized as follows. In Section 2, we review related studies in the literature of partner selection in cooperation networks. In Section 3, we build a patent cooperation network according to the participating organizations and the partnerships between them from the data collected from 2007 to 2015. Then, in Section 4, we analyze certain basic properties of the network and analyze three communities in the network after community detection. To predict the future partnerships between organizations in patent applications, we use link prediction approaches in Section 5 and compare their results with the data of cooperative patents in 2016. The last section is the discussion of this paper.

2. Literature Review
2.1. Partner Selection in Cooperation Networks

The existing studies on partner selection in cooperation networks mainly focused on the following three aspects: geographical locations of cooperators and the distance between them, the relevance of cooperator capabilities, and trust relationships between cooperators.

In the early years, geographical locations of cooperators and the distance between them played a crucial role in partner selection due to the undeveloped communication technology. Even now, they are still important factors in partner selection. Burhop et al. [17] and Drivas et al. [18] all found geographic distance was one of the most important obstacles for knowledge diffusion, including patent cooperation and transfer. Maggioni et al. [19,20] built knowledge networks to analyze the influence of geographical distance on innovation, and they realized that scientific and technological knowledge and patenting activities were both created and diffused through crucial nodes like universities, research institutions, and firms. Geerts et al. [21] analyzed the effect of the geographical dimension on the ambidexterity–performance relationship and analyzed the benefits of geographical proximity between technology exploitation and exploration for firms. By analyzing the existing literature and some cooperative data in a project, Gattringer et al. [22] found that geographical proximity contributed to the occurrence of cooperation and the exchange of knowledge. Mejdalani et al. [23] created a network of interregional co-patenting from 2000 to 2011 in Brazil and found that geography still played a fundamental role in their forming networks. Capaldo et al. [24] believed that geographical distance and organizational proximity were complementary and that organizational proximity could offset the far geographical distance to some extent in cooperation, and they suggested selecting partners at close range if the cooperative organizations worked in different domains.

In recent years, a growing number of studies have focused on the complementarity of cooperator capabilities. Shah et al. [25] proposed that the resources and capabilities of partners should be complementary rather than overlapped. Baum et al. [26] believed that knowl-
edge complementarity should play an important role in partner selection and designed a partner selection model by building an industry network to ensure that firms’ knowledge bases matched. Wei et al. [27] presented a two-stage partner selection framework according to the collaborative capabilities and sustainability of cooperators. Jee et al. [28] presented a patent-based framework to select suitable partners for technology-based firms according to their complementary capabilities, potential of learning, and risk from knowledge spillovers. Savin et al. [29] proposed an innovation network with endogenous absorptive capacity to indicate the cooperation between firms in different locations; they realized that partner selection between firms is driven by absorptive capacity that is influenced by distance and R&D investment allocation. Lv et al. [30] combined the capabilities of alternative partners, the potential conflicts between different partners, and the complementarity of their innovation sources to select partners in a supply chain collaboration.

Moreover, partner selection based on the trust relationships between cooperators is an important research aspect. Bierly et al. [31] analyzed the influence of trust among individuals and firms on the selection of an alliance partner. Campo et al. [32] used a structural equation model based on partial least squares to analyze the impact of a partner’s reputation and prior partnering experience on building trust, and the result they obtained is strong, positive, and significant. Reusen et al. [33] verified the influence of third-party information on trust relationships and studied how trust relationships affect the selection of partners. Chen et al. [34] proposed a dual-factor framework to evaluate and select cooperative partners by measuring collaborative level and individual level. In addition, some studies based on other factors have been proposed for partner selection [35,36].

2.2. Link Prediction in Complex Networks

Link prediction is often used to find the missing links in social networks and predict whether pairs of unconnected nodes will connect in the future based on the existing links. The existing link prediction methods are generally divided into three categories, namely similarity-based methods, probability-based methods, and learning-based methods.

Similarity-based methods measure the similarity of any pairs of unconnected nodes to predict whether the link should exist depending on the local and global information [37–39]. Zeng [40] combined the Common Neighbors (CN) index and the Preferential Attachment (PA) index to evaluate the likelihood of the existence of links. Probability-based methods evaluate the possibility of inexistent links by building statistical probability models [41,42]. Das et al. [43] proposed a probabilistic link prediction method based on a stochastic Markov model to predict links in time-varying networks. Learning-based methods use algorithms and intelligent frameworks for link prediction, such as metaheuristic ones [44] and machine learning [45–47]. Williamson [48] presented a Bayesian nonparametric approach on sparse networks that combined structure explanation and predictive performance for link prediction. Muniz et al. [49] combined contextual, temporal, and topological information to improve the link prediction accuracy. Because these link prediction methods can predict the links between unconnected nodes, they are suitable for partner selection in patent cooperation networks.

Some common link prediction approaches are introduced as follows, which are also used in this paper.

(1) Common Neighbors (CN) index [50]

The CN index can be used to evaluate whether an inexistent link between any pairs of unconnected nodes is a missing link or will exist in the future by counting their common neighbors. The CN index assumes that two unconnected nodes are more likely to connect if they have more common neighbors, and it is defined as Equation (1):

\[
S_{CN}(i,j) = |\Gamma(i) \cap \Gamma(j)|
\]  

where \(\Gamma(i)\) denotes the neighbor set of node \(i\).

(2) Adamic–Adar (AA) index [51]
The AA index is proposed to calculate a similarity score between two web pages depending on shared features. It can also be regarded as a variant of the CN index that assigns the common neighbors of two nodes higher weight if they have lower degrees. The AA index is defined as Equation (2):

\[
S_{AA}(i, j) = \sum_{k \in \Gamma(i) \cap \Gamma(j)} \frac{1}{\log |\Gamma(k)|}
\]  

(3) Resource Allocation (RA) index [52]

The RA index supposes that some resources can be sent from node \(i\) to another disconnected node \(j\) via their common neighbors, and the similarity between the two nodes can be calculated according to the resources that node \(j\) received. Similar to the AA index, the RA index is defined as Equation (3):

\[
S_{RA}(i, j) = \sum_{k \in \Gamma(i) \cap \Gamma(j)} \frac{1}{|\Gamma(k)|}
\]

(4) Leicht–Holme–Newman Local (LHNL) index [53]

The LHNL index supposes that two unconnected nodes are similar if their common neighbors are self-similar to themselves. It is defined as Equation (4):

\[
S_{LHNL}(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i)||\Gamma(j)|}
\]

(5) Local Naïve Bayes-Based Common Neighbors (LNBCN) index [54]

The LNBCN index is based on the naïve Bayes theory that different common neighbors of two nodes play different roles and have different contributions to the calculation of similarity. It is defined as Equation (5):

\[
S_{LNBCN} = \sum_{k \in \Gamma(i) \cap \Gamma(j)} \left[ \log \left( \frac{C_k}{1 - C_k} \right) + \log \left( \frac{1 - D}{D} \right) \right]
\]

where \(C_k\) is the clustering coefficient of node \(k\) and \(D\) is the network density.

(6) Random Walk with Restart (RWR) [55]

RWR is based on random walk; it assumes that some particles can walk randomly between any two nodes, and the importance of a link can be measured by counting the number of times that particles pass through it in a specified number of rounds. It is defined as Equation (6):

\[
S_{RWR}(i, j) = q_{ij} + q_{ji}
\]

where \(q_{ij}\) denotes the probability that particles walk randomly via real link \(E(i, j)\) from node \(i\) to node \(j\) in a specified number of steps.

(7) SimRank [56]

SimRank is a structural context similarity measurement and can be used to measure how soon two random walkers arrive at the same position from different nodes. It is a recursive function and can be defined as Equation (7):

\[
S_{SimRank}(i, j) = \begin{cases} \alpha \sum_{m=1}^{\infty} \frac{1}{\Gamma_m(i)} \sum_{m=1}^{\infty} S_{SimRank}(\Gamma_m(i), \Gamma_m(j)) \end{cases}, \quad i \neq j \\
1, \quad i = j
\]

where \(\alpha \in (0, 1)\) is a constant and \(\Gamma_m(i)\) denotes the \(m\)th neighbor of node \(i\).

(8) Knowledge-Dissemination-Based Link Prediction (KDLP) [57]
KDLP defines the knowledge quantity for each node by calculating its H-index and giving weights to the links in a network based on knowledge dissemination between nodes. It is defined as Equation (8):

$$S_{KDLP}(i, j) = \sum_{i=1}^{n} \beta^n (WA)^n_{ij}$$

where \(\beta\) is a free parameter, \(WA\) denotes the weighted adjacency matrix of a network, and \(n\) denotes the length of a path between node \(i\) and node \(j\).

3. Interorganizational Patent Cooperation Network

3.1. Data Sources

We collected the data of cooperative patents applied for by organizations in China between 2007 and 2015 from China National Intellectual Property Administration (https://www.cnipa.gov.cn). There are 46,492 cooperative patents contained in these data. Figure 1 shows the number of patent applications per year. As can be seen from Figure 1, the number of cooperative patents applied for by organizations in China shows an increasing trend year by year.

![Figure 1. The number of cooperative patents applied for by organizations in China per year from 2007 to 2015.](image)

In this paper, we treated all the data as a whole of historical information regardless of years. Considering that the time span of these data is as long as nine years, some applicants for different patents may be identical. After the deletion of duplicated items, there were 20,174 remaining patents, and the participating organizations in each patent were different from others. By observing these cooperative patents, we counted the number of participating organizations as 16,210 and found that they could be divided into three categories, namely university, company, and research institute, and the number of organizations in each category was 573, 13,740, and 1897.

- University, a high-level educational institution in which people study for degrees and engage in academic research. A university, whether public or private, is primarily designed to cultivate talents rather than make a profit. Compared with a company and a research institute, research in a university occupies the dominant position while the proportions of development and production are very small.
- Company, a business organization that earns profits by developing and selling goods or services. Different from a university, a company generally takes product development as the primary business and hence has a much stronger development capacity.
Currently, more and more companies are beginning to pay attention to research and set up affiliated research institutes dedicated to research work.

- Research institute is an establishment endowed for doing basic research or applied research. A research institute may be independent or affiliated to a university or a company. An independent research institute can conduct both research and development, but the research capacity is weaker than a university and the development capacity is weaker than a company. Affiliated research institutes usually engage in research related to their parent organizations as a supplement to the weakness. As a whole, a research institute does not pay absolute attention to research as a university or concentrate on development as a company, but it maintains a balance between research and development.

3.2. Network Building

In order to facilitate the analysis of the partnerships in patent cooperation between different organizations, we built an interorganizational patent cooperation network of 16,210 organizations and 23,146 partnerships between different organizations after removing duplicate partnerships between organizations from the collected data. In this unweighted and undirected network, a node corresponded to a participating organization and an existing edge between two nodes indicated that the corresponding organizations had applied for the same patent. For convenience, we used $N = (V, E)$ to represent the patent cooperation network, $V$ was the set of nodes, and $E$ was the set of edges. Moreover, we used $n$ for the number of nodes and $m$ for the number of edges, therefore, $n = 16,210$ and $m = 23,146$.

We represented the network graphically so that it could be observed visually. Figure 2 shows the visualization network diagram processed by the Gephi software (version 0.9.2, https://gephi.org/). In Figure 2, a node represents an organization and an edge connecting two nodes indicates a patent partnership between two corresponding organizations in patent. Moreover, nodes and edges of the same color show that they belong to the same community, and the connections between nodes within the same community are closer than those between nodes in different communities.

![Figure 2. The visualization of the interorganizational patent cooperation network from 2007 to 2015.](image)

4. Analysis of the Interorganizational Patent Cooperation Network

4.1. Basic Properties

We measured certain basic properties of the network for analysis, including average degree of nodes, network diameter and density, average clustering coefficient, average path length, betweenness centrality (BC), closeness centrality (CC), and eigenvector centrality (EC) of nodes.
4.1.1. Average Degree of Nodes

Considering that the patent cooperation network is an unweighted and undirected network, we calculated the average degree of nodes $\bar{d}$ by Equation (9) according to the number of nodes and edges in the network:

$$\bar{d} = \frac{\sum_{i=1}^{|V|} d_i}{n} = \frac{2m}{n}$$  \hspace{1cm} (9)

where $d_i$ denotes the degree of node $i$ in the network and it shows the number of partnerships between organization $i$ and others.

After calculation, the average degree of nodes in the network is 2.856, meaning that each node has approximately three partners on average. However, only a few nodes in the network have a large degree, while most nodes show a small degree.

It is worth noting that the degree of a node does not represent the number of patents that the organization participated in; it only indicates the number of different partners with which the organization had cooperated. Further statistics include the sum of degrees for the top 25% nodes that accounts for 74.46% of the sum of degrees for all the nodes (34,150/46,296). Obviously, the degree of nodes in the network conforms to the Pareto principle.

Moreover, universities account for an overwhelming majority of seats, i.e., 38 of the top 40. The reason is that, on the one hand, cooperating with universities can improve the success rate for patent applications, especially top universities in China with the strongest research capacity. On the other hand, research may be more important than development and production in patent applications.

4.1.2. Network Diameter and Network Density

Network diameter refers to the maximum path distance between any two nodes in the network, which is generally measured by the number of edges. Because the network is not a completely connected graph and contains 710 weakly connected components, we set its network diameter as the largest diameter in all connected subgraphs, with a value of 14.

Network density is the ratio of the number of edges to all the possible links between nodes in a network. A higher network density means more edges exist in the network and corresponds to a denser graph. On the contrary, a lower network density means less edges exist in the network and corresponds to a sparser graph. Network density $D$ can be calculated by Equation (10):

$$D = \frac{2m}{n(n-1)}$$  \hspace{1cm} (10)

The density of this interorganizational patent cooperation network is 0.0002, indicating that the network is a sparse graph.

4.1.3. Average Clustering Coefficient

Clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. For a node $i$ in the network, its clustering coefficient $C_i$ can be calculated by Equation (11):

$$C_i = \frac{2\sum_{|e_{ij}|=1} |e_{jk}|}{d_i(d_i-1)}$$  \hspace{1cm} (11)

where $e_{ij}$ denotes the edge between node $i$ and node $j$, $|e_{ij}| = 1$ if $e_{ij}$ exists in the network, otherwise, $|e_{ij}| = 0$. $d_i$ denotes the degree of node $i$.

The average clustering coefficient of the network $\bar{C}$ can be obtained by Equation (12):

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i$$  \hspace{1cm} (12)
For the interorganizational patent cooperation network, the average clustering coefficient is 0.637. It is significantly higher than the random network generated by the same set of nodes.

4.1.4. Average Path Length

Average path length is defined as the average length of the shortest paths for all possible pairs of network nodes. Average path length of a network $\mathcal{L}$ can be calculated by Equation (13):

$$\mathcal{L} = \frac{\sum_{i \neq j} L_{ij}}{n(n-1)}$$

(13)

where $L_{ij}$ is the shortest path length between node $i$ and node $j$.

In this network, the average path length $\mathcal{L} = 4.626$ is less than $\ln n = 9.693$. Combined with the average clustering coefficient of the network, it can be determined that the patent cooperation network conforms to the characteristics of a small-world network.

4.1.5. Betweenness Centrality, Closeness Centrality, and Eigenvector Centrality

Betweenness centrality of a node refers to the proportion of shortest paths between any other nodes that contains the node. Betweenness centrality of node $i$, $BC_i$, can be calculated by Equation (14):

$$BC_i = \sum_{j=1}^{n} \sum_{k=1}^{n} I_{jk}(i), \quad i \neq j \neq k, \ j < k$$

(14)

where $I_{jk}(i)$ is an indicative function that indicates whether node $i$ is located on the shortest path between node $j$ and node $k$.

Closeness centrality of a node is based on the average shortest path length between the node and others. A smaller average shortest path length corresponds to a greater closeness centrality. Closeness centrality of node $i$, $CC_i$, can be calculated by Equation (15):

$$CC_i = \frac{n - 1}{\sum_{j \neq i} L_{ij}}$$

(15)

Eigenvector centrality of a node is based on the importance of neighbors of the node. Eigenvector centrality of node $i$, $EC_i$, can be calculated by Equation (16):

$$EC_i = c \sum_{j=1}^{n} a_{ij}x_j$$

(16)

where $a_{ij}$ indicates the relationship between node $i$ and node $j$, if $e_{ij} \in E, a_{ij} = 1$, otherwise, $a_{ij} = 0$, $c$ is a constant which can be determined by $x = cAx$, and $A$ is the adjacent matrix of $N$.

Taking Tsinghua University as an example, its betweenness centrality is 11,025,318.2101 and eigenvector centrality is 0.8572, they are the second largest in all the organizations, second only to State Grid Corporation of China at 24,033,631.0808 and 1.0000. In addition, the closeness centrality of Tsinghua University is 0.3329.

Table 1 lists the top-30 largest degree organizations and their betweenness centralities (BC), closeness centralities (CC), and eigenvector centralities (EC) in the interorganizational patent cooperation network we built.
Table 1. The top-30 largest degree organizations and their betweenness centrality (BC), closeness centrality (CC), and eigenvector centrality (EC) in the interorganizational patent cooperation network.

| Ranking | Name                                      | Degree | BC (1)       | CC (2)      | EC (3)      |
|---------|-------------------------------------------|--------|--------------|-------------|-------------|
| 1       | Tsinghua University                       | 662    | 11,025,318.2101 | 0.3329      | 0.8572      |
| 2       | Zhejiang University                      | 574    | 9,544,670.7759  | 0.3039      | 0.6333      |
| 3       | State Grid Corporation of China           | 531    | 24,033,631.0808 | 0.3825      | 1.0000      |
| 4       | Shanghai Jiaotong University              | 445    | 8,431,466.6655  | 0.3331      | 0.5667      |
| 5       | South China University of Technology      | 399    | 6,694,671.3031  | 0.3175      | 0.4001      |
| 6       | East China University of Science and Technology | 330    | 5,348,480.3721  | 0.3165      | 0.3018      |
| 7       | Southeast University                      | 299    | 5,976,122.7570  | 0.3180      | 0.3350      |
| 8       | Jiangnan University                      | 273    | 3,861,144.9151  | 0.2643      | 0.1294      |
| 9       | Donghua University                       | 263    | 3,903,748.7569  | 0.3006      | 0.1809      |
| 10      | Tianjin University                       | 252    | 3,800,105.5266  | 0.3233      | 0.3916      |
| 11      | Huazhong University of Science and Technology | 252    | 3,094,261.8730  | 0.3082      | 0.3192      |
| 12      | Xi’an Jiaotong University                 | 239    | 3,034,869.6657  | 0.3097      | 0.3180      |
| 13      | Central South University                  | 222    | 3,090,352.3719  | 0.2810      | 0.1273      |
| 14      | Chongqing University                     | 215    | 2,649,603.8271  | 0.3052      | 0.2456      |
| 15      | North China Electric Power University    | 209    | 1,170,028.5469  | 0.3017      | 0.3801      |
| 16      | Dalian University of Technology           | 205    | 3,726,326.0807  | 0.3178      | 0.2669      |
| 17      | Zhejiang University of Technology        | 204    | 3,000,924.7461  | 0.3071      | 0.1656      |
| 18      | Wuhan University                         | 202    | 2,458,444.0919  | 0.3105      | 0.3245      |
| 19      | Sun Yat-Sen University                    | 197    | 3,656,003.5214  | 0.2884      | 0.1432      |
| 20      | Sichuan University                       | 195    | 3,439,624.5204  | 0.3052      | 0.2114      |
| 21      | University of Science and Technology Beijing | 192    | 2,867,702.3456  | 0.3042      | 0.1824      |
| 22      | Tongji University                        | 189    | 2,613,293.6225  | 0.3074      | 0.1821      |
| 23      | Beijing University of Chemical Technology | 177    | 3,052,419.6455  | 0.3020      | 0.1533      |
| 24      | Shandong University                      | 165    | 2,401,462.403   | 0.3100      | 0.2414      |
| 25      | Nanjing Tech University                  | 158    | 2,378,152.4567  | 0.2993      | 0.1282      |
| 26      | Beihang University                       | 155    | 2,400,896.3232  | 0.3080      | 0.1560      |
| 27      | Peking University                        | 149    | 2,520,874.4966  | 0.3099      | 0.1821      |
| 28      | Shanghai University                      | 147    | 1,873,093.8985  | 0.2781      | 0.0902      |
| 29      | Northeastern University                   | 141    | 1,628,495.7353  | 0.2916      | 0.1133      |
| 30      | Fudan University                         | 140    | 2,171,318.4669  | 0.2887      | 0.1032      |

4.2. Community Detection and Analysis

Due to the large size of the network, it was not enough to analyze the network as a whole. In addition to the analysis of the above basic properties, we further analyzed some locals of the network by community detection.

The whole network could be divided into 760 communities and the corresponding network modularity was 0.809. It was difficult to analyze all the communities due to their large number; therefore, three communities were selected as representatives and each of them contained the organization with the largest degree in each category.

4.2.1. The Community Containing the University with the Largest Degree

Tsinghua University has the largest degree of all the universities, even of all the participating organizations. Therefore, the community that contains Tsinghua University is the primary analysis object. Figure 3 shows the structure of this community in the network, the nodes in this community are shown in red, and Tsinghua University owns the largest size in the figure.

There are 488 organizations in this community, the majority of which are companies. One possible reason is that Tsinghua University, as one of the top universities in China, has an excellent research capacity, which has allowed it to attract more cooperate interests from companies. In addition, the degree of most nodes in this community is 1 (see the lower right in Figure 3), and they only have partnerships with Tsinghua University.
Table 1. The top-30 largest degree organizations and their eigenvector centrality (EC), betweenness centrality (BC), and closeness centrality (CC).

| Rank | Name                                | Degree | EC          | BC           | CC           |
|------|-------------------------------------|--------|-------------|--------------|--------------|
| 1    | Tsinghua University                 | 1109   | 0.3329      | 0.8572       |              |
| 2    | Zhejiang University                 | 574    | 0.3309      | 0.6333       |              |
| 3    | State Grid Corporation of China     | 531    | 0.3825      | 1.0000       |              |
| 4    | Shanghai Jiaotong University        | 445    | 0.3331      | 0.5667       |              |
| 5    | South China University of Technology| 399    | 0.3175      | 0.4001       |              |
| 6    | East China University of Science and Technology | 330 | 0.3165 | 0.3018 |
| 7    | Southeast University                 | 299    | 0.3180      |              | 0.3350       |
| 8    | Jiangnan University                 | 273    | 0.2643      |              | 0.1294       |
| 9    | Donghua University                  | 263    | 0.3006      |              | 0.1809       |
| 10   | Central South University            | 222    | 0.2810      |              | 0.1273       |
| 11   | Xi'an Jiaotong University           | 239    | 0.3097      |              | 0.3180       |
| 12   | Huazhong University of Science and Technology | 252 | 0.3082 | 0.3192 |
| 13   | Chongqing University                | 215    | 0.3052      |              | 0.2456       |
| 14   | North China Electric Power University | 209 | 0.3017  | 0.3801 |
| 15   | University of Science and Technology Beijing | 192 | 0.3042 | 0.1824 |
| 16   | Zhejiang University of Technology   | 204    | 0.3071      |              | 0.1656       |
| 17   | Wuhan University                    | 202    | 0.3105      |              | 0.3245       |
| 18   | Nanjing Tech University             | 158    | 0.2993      |              | 0.1282       |
| 19   | Shandong University                 | 165    | 0.3100      |              | 0.2414       |
| 20   | Beihang University                  | 155    | 0.3080      |              | 0.1560       |
| 21   | Shanghai University                 | 147    | 0.2781      |              | 0.0902       |
| 22   | Fudan University                    | 140    | 0.2887      |              | 0.1032       |

Figure 3. The structure of the community that contains Tsinghua University.

Combined with observations of other communities that include universities with a large degree, we can find that companies are prone to cooperate with universities. It is probably because the patent cooperation between companies and universities can complement their advantages and help to distribute their benefits.

4.2.2. The Community Containing the Company with the Largest Degree

The second community we analyzed contains State Grid Corporation of China, which has the largest degree of any of the companies. Figure 4 shows the structure of this community in the network, and the nodes in the community are shown in green. Similar to Tsinghua University in Figure 3, we made the node that represents State Grid Corporation of China larger than other nodes in Figure 4 for observation.

Figure 4. The structure of the community that contains State Grid Corporation of China.

This community is the largest of all, which contains 1109 organizations in the three categories. In addition to comprehensive universities, the universities in partnership with State Grid Corporation of China also include some specialized universities in electric power, such as Northeast Electric Power University and North China Electric Power University.
4.2.3. The Community Containing the Research Institute with the Largest Degree

In the third community we analyzed, the Institute of Process Engineering, Chinese Academy of Sciences, had the largest degree of any of the research institutes. Figure 5 shows the structure of this community, and the nodes in the community are shown in saffron yellow. Similarly, we made the node that represents the Institute of Process Engineering, Chinese Academy of Sciences, larger than other nodes in this community.

As can be seen from Figure 5, this community contains 443 organizations, and it can be further divided into two small communities. Although the Institute of Process Engineering, Chinese Academy of Sciences, is not the largest degree node in this community, it is the absolute center of the small community on the left.

According to the structure of the above communities, small- and medium-sized companies prefer to select universities and research institutes as partners to cooperate, especially a famous university. For this reason, the university has a strong scientific research capacity, which a company lacks. On the contrary, the company owns a strong production capacity, and that is exactly what a university lacks. The complementarity of capabilities contributes to the development of patents because patents require both theoretical study and practical application.

5. Prediction of Future Relationships between Organizations

By analyzing certain basic properties and the three communities in the interorganizational patent cooperation network, we found that the network had some typical characteristics of social networks. In this section, we try to predict future partnerships between these organizations based on the patent cooperation network by introducing link prediction approaches, which are widely applied to predict the relationships between nodes in social networks.

The link prediction approaches we used in this paper were mentioned in Section 2.2, namely the CN index, the AA index, the RA index, the LHNL index, the LNBCB index, RWR, SimRank, and KDLP.

In order to verify the effectiveness of the above link prediction approaches on the partnership prediction between organizations in a patent cooperation network, we also collected the data of patents applied for in 2016 from China National Intellectual Property Administration as a validation.

The data from patent applications in 2016 contained 10,367 patents. After using a processing similar to the one adopted for the data from 2007 to 2015, 5555 partnerships were reserved. We ran the above eight link prediction approaches on the 2007–2015 patent
cooperation network and compared their prediction results with the data of 2016. In addition, we selected precision as the evaluation indicator to calculate the prediction accuracy of the link prediction approaches on the 2016 patent cooperation network. Precision is defined as Equation (17):

\[
\text{Precision} = \frac{n}{N'}
\]

where \(n\) denotes the number of partnerships that are both in the prediction result of a link prediction approach and in the edge set of the 2016 patent cooperation network. \(N'\) is the number of edges in the 2016 patent cooperation network.

Figure 6 shows the precision results of the above eight link prediction approaches.

![Figure 6. Precision results of the eight link prediction approaches.](image)

According to the prediction results shown in Figure 6, all the above eight link prediction approaches can obtain a precision value greater than 0.7. All the precision results of these link prediction approaches can be divided into three levels as follows: the CN index obtains the highest prediction accuracy and is much better than other link prediction approaches; the AA index and RWR obtain the second-best level of prediction accuracy; the RA, LHNL, LNBN, SimRank, and KDLP approaches obtain similar prediction accuracies, but their results are lower than the CN index, the AA index and RWR and represent the third-best level of prediction accuracy.

The CN index performs better than the others and also shows that the China interorganizational patent cooperation network is a social network. Just as people, an organization prefers to choose the partners of its partners rather than strangers. Moreover, having more common partners between two organizations that have never worked together means that a new partnership is more likely to emerge.

Although the prediction accuracies obtained by these link prediction approaches are mixed, they are all acceptable. Therefore, link prediction approaches are effective in the prediction of partnerships between organizations, and they can be applied to predict future partnerships based on previous partnerships in patent application.

6. Discussion

By collecting the cooperative patents applied for by organizations in China from 2007 to 2015, we built a patent cooperation network for this paper, including 16,210 nodes and 20,174 edges. In this network, a node represents an organization, which could be a university, a company, or a research institute. An edge between a pair of connected nodes indicates that there was a partnership between the two organizations. After analyzing
certain basic properties of the patent cooperation network, such as average degree of nodes, network diameter and network density, average clustering coefficient, average path length, betweenness centrality, closeness centrality, and eigenvector centrality of each node, we found that the network was in conformity with the Pareto principle and a small-world network and hence it can be regarded as a social network. Then, because it was difficult to analyze the network as a whole due to its large scale, we divided the network into 760 communities and selected three of them as examples to illustrate our study. Each of the three communities contained the organization with the largest degree in each category.

In order to select suitable partners for organizations in future patent cooperation, we introduced eight link prediction approaches that are commonly used in social networks to predict the possible future partnerships between organizations in the China interorganizational patent cooperation network. Moreover, we collected the data of interorganizational patent cooperation in 2016 and built another network to evaluate the prediction accuracy of these link prediction approaches using the precision metric. The results show all the eight link prediction approaches can obtain more than 70% prediction accuracy. They demonstrate the effectiveness of the link prediction approaches in partner selection for organizations in patent cooperation and show that the trust transitivity among cooperators based on historical cooperation is also an important factor to be considered when an organization makes partner selection.

Moreover, we processed each partnership depending on the location of the cooperators and found that the cooperators in 86.7% of the partnerships are in the same province. It shows that geographical location and distance are still important considerations in partner selection for most organizations. Then, we divided the interorganizational patent cooperation network into communities for local analysis and found that companies prefer to cooperate with universities and research institutes according to the results of community detection. This indicates that the complementarity of capabilities also plays an important role in partner selection for organizations.

The following conclusions can be made from this paper. (1) Because the inner core of organizations is people, the interorganizational patent cooperation network is essentially a social network. By analyzing certain basic properties of the patent cooperation network, we found that it conforms to the Pareto principle and a small-world network. Therefore, link prediction approaches commonly used in social networks can be introduced to predict the possible future partnerships between organizations in this network. (2) Through the analysis of historical partnerships, the communities, and the results of link prediction approaches, we realize that geographical location, the complementarity of capabilities, and the trust transitivity based on historical cooperation are factors that are worthy of consideration in partner selection for organizations. Although the trust transitivity based on historical information has a positive influence on partner selection, consideration of all three factors will bring a better guidance to organizations.

Based on our work in this paper, we can make the following recommendations for organizations when selecting partners. For organizations that have never worked before with others, geographical location and the complementarity of capabilities need to be considered in the first partner selection. Using the example of small- and medium-sized companies, universities and research institutes located closer to them are ideal initial partners. For organizations that have already worked together, historical cooperation information also needs to be considered. Link prediction approaches, such as the CN index, can play a good guiding role in finding new partners.

In the future, we plan to carry out further research on two aspects. On the one hand, we will analyze the influence of capacity complementarity among different categories of organizations and combine the capacities of organizations and historical cooperation information for partner selection in patent cooperation. On the other hand, we will deploy more advanced link prediction approaches on the patent cooperation network to further improve the prediction accuracy.
**Author Contributions:** Conceptualization, H.Q. and K.C.; methodology, H.Q.; software, K.C.; validation, H.Q., W.C., and K.C.; formal analysis, K.C.; investigation, H.Q.; resources, H.Q.; data curation, H.Q.; writing—original draft preparation, H.Q.; writing—review and editing, W.C.; visualization, K.C.; supervision, W.C.; project administration, W.C.; funding acquisition, W.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Chinese National Funding of Social Sciences, grant number 19BGL017; the Natural Science Foundation of Heilongjiang Province, grant number LH2019GC03; the Social Science Foundation of Heilongjiang Province, grant number 19GLC160; and the Fundamental Research Funds for the Central Universities, grant numbers 3072020CFW0915 and 3072020CFW9010.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the data content is in Chinese.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Delerue, H. Shadow of joint patents: Intellectual property rights sharing by SMEs in contractual R&D alliances. *J. Bus. Res.* 2018, 87, 12–23. [CrossRef]

2. Geum, Y.; Lee, S.; Yoon, B.; Park, Y. Identifying and evaluating strategic partners for collaborative R&D: Index-based approach using patents and publications. *Technovation* 2013, 33, 211–224. [CrossRef]

3. Song, B.; Seol, H.; Park, Y. A patent portfolio-based approach for assessing potential R&D partners: An application of the Shapley value. *Technol. Forecast. Soc. Chang.* 2016, 103, 156–165. [CrossRef]

4. Sun, H.; Geng, Y.; Hu, L. Measuring China’s new energy vehicle patents: A social network analysis approach. *Energy* 2018, 153, 685–693. [CrossRef]

5. Arranz, N.; de Arroyabe, J.C.F. The choice of partners in R&D cooperation: An empirical analysis of Spanish firms. *Technovation* 2008, 28, 88–100. [CrossRef]

6. Gong, B.; Yang, C.-L. Cooperation through indirect reciprocity: The impact of higher-order history. *Games Econ. Behav.* 2019, 118, 316–341. [CrossRef]

7. Liu, C.; Wang, J.; Li, X. The link weight adjustment considering historical strategy promotes the cooperation in the spatial prisoner’s dilemma game. *Phys. A* 2020, 554, 124691. [CrossRef]

8. Alcañiz, I. Partner selection in international environmental networks: The effect of skills and money on cooperation in the Global South. *Environ. Sci. Policy* 2016, 55, 107–115. [CrossRef]

9. Han, J.; Teng, X.; Cai, X. A novel network optimization partner selection method based on collaborative and knowledge networks. *Inf. Sci.* 2019, 484, 269–285. [CrossRef]

10. Knoben, J.; Gilsing, V.A.; Krijkamp, A.R. From homophily through embeddedness to strategy: The role of network accuracy in partner selection choices. *Long Range Plan.* 2019, 52, 86–102. [CrossRef]

11. Mladineo, M.; Celar, S.; Celent, L.; Crnjac, M. Selecting manufacturing partners in push and pull-type smart collaborative networks. *Adv. Eng. Inform.* 2018, 38, 291–305. [CrossRef]

12. Lejdel, B. Negotiation and cooperation between agents for generalizing geographic objects. *J. Comput. Lang.* 2019, 51, 15–27. [CrossRef]

13. Kölle, F. Heterogeneity and cooperation: The role of capability and valuation on public goods provision. *J. Econ. Behav. Organ.* 2015, 109, 120–134. [CrossRef] [PubMed]

14. Wu, J. Cooperation with competitors and product innovation: Moderating effects of technological capability and alliances with universities. *Ind. Mark. Manag.* 2014, 43, 199–209. [CrossRef]

15. Bravo, G.; Squazzoni, F.; Boero, R. Trust and partner selection in social networks: An experimentally grounded model. *Soc. Networks* 2012, 34, 481–492. [CrossRef]

16. Benitez-Acila, C.; Hartmann, A.; Devulfs, G. Interplay of relational and contractual governance in public-private partial-partnerships: The mediating role of relational norms, trust and partners’ contribution. *Int. J. Proj. Manag.* 2018, 36, 429–443. [CrossRef]

17. Burhop, C.; Wolf, N. The German market for patents during the “second industrialization”, 1884–1913: A gravity approach. *Bus. Hist. Rev.* 2013, 87, 69–93. [CrossRef]

18. Drivas, K.; Economidou, C. Is geographic nearness important for trading ideas? Evidence from the US. *J. Technol. Transf.* 2015, 40, 629–662. [CrossRef]

19. Maggoni, M.A.; Nosvelli, M.; Uberti, T.E. Space versus networks in the geography of innovation: A European analysis. *Pap. Reg. Sci.* 2007, 86, 471–493. [CrossRef]

20. Maggoni, M.A.; Uberti, T.E. Knowledge networks across Europe: Which distance matters? *Ann. Reg. Sci.* 2008, 43, 691–720. [CrossRef]
21. Geerts, A.; Leten, B.; Belderbos, R.; Van Looy, B. Does Spatial Ambidexterity Pay Off? On the Benefits of Geographic Proximity Between Technology Exploitation and Exploration. *J. Prod. Innov. Manag.* 2017, 35, 151–163. [CrossRef]
22. Gattringer, R.; Wiener, M.; Strehl, F. The challenge of partner selection in collaborative foresight projects. *Technol. Forecast. Soc. Chang.* 2017, 120, 298–310. [CrossRef]
23. Mojdalani, A.; Gonçalves, E.; Rodriguez, R.S. Why cooperate? The determinants of forming inter-regional connections in the Brazilian patent network. *Economía* 2020. [CrossRef]
24. Capaldo, A.; Petruzzielli, A.M. Partner Geographic and Organizational Proximity and the Innovative Performance of Knowledge-Creating Alliances. *Eur. Manag. Rev.* 2014, 11, 63–84. [CrossRef]
25. Shah, R.H.; Swaminathan, V. Factors influencing partner selection in strategic alliances: The moderating role of alliance con-text. *Strateg. Manag. J.* 2008, 29, 471–494. [CrossRef]
26. Baum, J.A.; Cowan, R.; Jonard, N. Network-Independent Partner Selection and the Evolution of Innovation Networks. *Manag. Sci.* 2010, 56, 2094–2110. [CrossRef]
27. Wei, F.; Feng, N.; Yang, S. A conceptual framework of two-stage partner selection in platform-based innovation ecosystems for servitization. *J. Clean. Prod.* 2020, 262, 121431. [CrossRef]
28. Jee, S.J.; Sohn, S.Y. Patent-based framework for assisting entrepreneurial firms’ R&D partner selection: Leveraging their limited resources and managing the tension between learning and protection. *J. Eng. Technol. Manag.* 2020, 57, 101575.
29. Savin, I.; Egbertokun, A. Emergence of innovation networks from R&D cooperation with endogenous absorptive capacity. *J. Econ. Dyn. Control.* 2016, 64, 82–103.
30. Lv, B.; Qi, X. Research on partner combination selection of the supply chain collaborative product innovation based on product innovative resources. *Comput. Ind. Eng.* 2019, 128, 245–253. [CrossRef]
31. Bierly, P.E.; Gallagher, S. Explaining Alliance Partner Selection: Fit, Trust and Strategic Expediency. *Long Range Plan.* 2007, 40, 134–153. [CrossRef]
32. Campo, J.D.S.P.G.D.; Pardo, I.P.G.; Perlines, F.H. Influence factors of trust building in cooperation agreements. *J. Bus. Res.* 2014, 67, 710–714. [CrossRef]
33. Reusen, E.; Stouthuysen, K. Trust transfer and partner selection in interfirm relationships. *Account. Organ. Soc.* 2020, 81, 101081. [CrossRef]
34. Chen, W.; Goh, M. Mechanism for cooperative partner selection: Dual-factor theory perspective. *Comput. Ind. Eng.* 2019, 128, 254–263. [CrossRef]
35. Angué, K.; Ayerbe, C.; Mitkova, L. A method using two dimensions of the patent classification for measuring the technological proximity: An application in identifying a potential R&D partner in biotechnology. *J. Technol. Transf.* 2013, 39, 716–747. [CrossRef]
36. Petruzzielli, A.M.; Murgia, G. University–Industry collaborations and international knowledge spillovers: A joint-patent investigation. *J. Technol. Transf.* 2020, 45, 958–983. [CrossRef]
37. Xu, M.; Yin, Y. A similarity index algorithm for link prediction. In Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, China, 24–26 November 2017.
38. Sun, Q.; Hu, R.; Yang, Z.; Yao, Y.; Yang, F. An improved link prediction algorithm based on degrees and similarities of nodes. In Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 24–26 May 2017.
39. Liu, Y.; Zhao, C.; Wang, X.; Huang, Q.; Zhang, X.; Yi, D. The degree-related clustering coefficient and its application to link prediction. *Phys. A Stat. Mech. its Appl.* 2016, 454, 24–33. [CrossRef]
40. Zeng, S. Link prediction based on local information considering preferential attachment. *Phys. A Stat. Mech. Appl.* 2016, 443, 537–542. [CrossRef]
41. Yang, Q.; Dong, E.; Xie, Z. Link prediction via nonnegative matrix factorization enhanced by blocks information. In Proceedings of the 2014 10th International Conference on Natural Computation (ICNC), Xiamen, China, 19–21 August 2014.
42. Chi, K.; Yin, G.; Dong, Y.; Dong, H. Link prediction in dynamic networks based on the attraction force between nodes. *Knowl.-Based Syst.* 2019, 181, 104792. [CrossRef]
43. Das, S.; Das, S.K. A probabilistic link prediction model in time-varying social networks. In Proceedings of the 2017 IEEE Interna-tional Conference on Communication, Paris, France, 21–25 May 2017.
44. Bastami, E.; Mahabadi, A.; Taghizadeh, E. A gravitation-based link prediction approach in social networks. *Swarm Evol. Comput.* 2019, 44, 176–186. [CrossRef]
45. Li, T.; Wang, B.; Jiang, Y.; Zhang, Y.; Yan, Y. Restricted Boltzmann Machine-Based Approaches for Link Prediction in Dynamic Networks. *IEEE Access* 2018, 6, 29940–29951. [CrossRef]
46. Zhang, C.; Zhang, H.; Yuan, D.; Zhang, M. Deep Learning Based Link Prediction with Social Pattern and External Attribute Knowledge in Bibliographic Networks. In Proceedings of the 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Chengdu, China, 15–18 December 2016.
47. Mohan, A.; Venkatesan, R.; Pramod, K. A scalable method for link prediction in large real world networks. *J. Parallel Distrib. Comput.* 2017, 109, 89–101. [CrossRef]
48. Williamson, S.A. Nonparametric network models for link prediction. *J. Mach. Learn. Res.* 2016, 17, 1–21.
49. Muniz, C.P.; Goldschmidt, R.R.; Choren, R. Combining contextual, temporal and topological information for unsupervised link prediction in social networks. *Knowl. Based Syst.* 2018, 156, 129–137. [CrossRef]

50. Lorrain, F.; White, H.C. Structural equivalence of individuals in social networks. *Soc. Netw.* 1977, 1, 67–98. [CrossRef]

51. Adamic, L.A.; Adar, E. Friends and neighbors on the Web. *Soc. Netw.* 2003, 25, 211–230. [CrossRef]

52. Zhou, T.; Lü, L.; Zhang, Y.-C. Predicting missing links via local information. *Eur. Phys. J. B* 2009, 71, 623–630. [CrossRef]

53. Leicht, E.; Holme, P.; Newman, M.E.J. Vertex similarity in networks. *Phys. Rev. E* 2006, 73, 026120. [CrossRef]

54. Liu, Z.; Zhang, Q.-M.; Lü, L.; Zhou, T. Link prediction in complex networks: A local naive Bayes model. *Europhys. Lett.* 2011, 96, 48007. [CrossRef]

55. Tong, H.; Faloutsos, C.; Pan, J.-Y. Fast Random Walk with Restart and Its Applications. In Proceedings of the Sixth International Conference on Data Mining (ICDM’06), Hong Kong, China, 18–22 December 2006.

56. Jeh, G.; Widom, J. Simrank: A measure of structural-context similarity. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, AB, Canada, 23–26 July 2002.

57. Zhou, W.; Jia, Y. Predicting links based on knowledge dissemination in complex network. *Phys. A Stat. Mech. Appl.* 2017, 471, 561–568. [CrossRef]